Exploiting empirical knowledge for automatic delineation of city centres from large-scale topographic databases

Lüscher, Patrick ; Weibel, Robert

Abstract: Current topographic databases rarely represent higher order geographic phenomena, such as city centres. However, such concepts are often referred to by humans and used in various forms of spatial analysis. Hence, the value and usability of topographic databases can greatly be improved by methods that automatically create such higher order phenomena through cartographic pattern recognition techniques, departing from the very detailed, geometry-oriented representations of topographic databases. As many higher order phenomena are only vaguely defined, this paper develops and evaluates a methodology to acquire definitional knowledge about geographic phenomena by participant experiments and use this knowledge to drive the cartographic pattern recognition process. The method is applied to acquire knowledge about British city centres and delineate referents of city centre from topographic data. City centres produced for ten British cities are compared to areas derived from alternative sources.

DOI: https://doi.org/10.1016/j.compenvurbsys.2012.07.001

Originally published at:
Lüscher, Patrick; Weibel, Robert (2013). Exploiting empirical knowledge for automatic delineation of city centres from large-scale topographic databases. Computers, Environment and Urban Systems, 37:18-34.
DOI: https://doi.org/10.1016/j.compenvurbsys.2012.07.001

*Title Page

Exploiting empirical knowledge for automatic delineation of city centres from large-scale topographic databases

Patrick Lüscher *, Robert Weibel

Department of Geography, University of Zurich, Winterthurerstrasse 190, 8057 Zurich (Switzerland)

* Corresponding author. Tel.: +41 58 319 26 23.

E-Mail addresses: patrick.luescher@geo.uzh.ch (P. Lüscher), robert.weibel@geo.uzh.ch (R. Weibel)

Research highlights

- Comprehensive methodology including semantic modelling, pattern recognition & evaluation
- Participant experiment to acquire definitional knowledge of UK city centres
- Computational implementation derives city centre typicality from topographic data
- High plausibility of generated city centre boundaries by comparison to other sources
- Good correspondence of computed typicality values to human assessment of images

Exploiting empirical knowledge for automatic delineation of city centres from large-scale topographic databases

Keywords: Pattern recognition, topographic database, semantic modelling, urban structure, city centre

Abstract

Current topographic databases rarely represent higher order geographic phenomena, such as city centres. However, such concepts are often referred to by humans and used in various forms of spatial analysis. Hence, the value and usability of topographic databases can greatly be improved by methods that automatically create such higher order phenomena through cartographic pattern recognition techniques, departing from the very detailed, geometry-oriented representations of topographic databases. As many higher order phenomena are only vaguely defined, this paper develops and evaluates a methodology to acquire definitional knowledge about geographic phenomena by participant experiments and use this knowledge to drive the cartographic pattern recognition process. The method is applied to acquire knowledge about British city centres and delineate referents of city centre from topographic data. City centres produced for ten British cities are compared to areas derived from alternative sources. F_1-scores between 0.45 and 0.88 are achieved, suggesting that the delineation produced plausible city centre areas. The benefits of our work are better (and user-driven) descriptions of complex geographic phenomena that can form the basis for accurately enriching topographic databases with additional semantics, thus yielding added value for the data producer and the end user.

1 Introduction

National Mapping Agencies (NMAs) and other data producers maintain and disseminate topographic datasets at the very fine scale. Designed as general purpose products, these datasets offer a wealth of (mainly geometric) information about individual objects. However, they do not model the higher order geographic phenomena required by many applications. For example, they model buildings and parking spaces, but not hospital complexes, districts and settlements (Chaudhry & Mackaness, 2008a; Chaudhry, Mackaness, & Regnauld 2009; Lüscher, Weibel, & Burghardt, 2009); they model height fields, but not the extent of hills, valleys and mountain ranges (Chaudhry & Mackaness, 2008b; Straumann, 2010).

Improving their datasets by providing more of such higher level semantics could help NMAs and other data producers to establish a more user-driven access to geographic information (Hart & Greenwood, 2003; Davies, Wood, & Fountain, 2005). This allows representing geographic space more closely to the way it is conceptualised by people, linking to the ideas of naïve geography (Egenhofer & Mark, 1995). Human spatial reasoning is chiefly qualitative, i.e. based on spatial relations and regions (Egenhofer & Mark, 1995; Montello, 2003). Representing geographic regions is thus beneficiary for many applications such as geographic information retrieval, navigation, and building gazetteers (Heinzle, Kopczynski, & Sester, 2003; Purves et al., 2007; Montello, 2003). For example, people might be interested in answers to queries such as "Where are city centre hotels?" Furthermore, having higher order phenomena in the database allows NMAs to respond better to customer requirements. Professionals of various disciplines maintain that concepts related to urban area and place, such as settlement, neighbourhood, townscape, and urban structure, are key spatial concepts (Davies, Holt, Green, Harding, & Diamond, 2009). This is often reflected in medium scale maps and maps for urban planning which emphasise urban structure (Steiniger, Lange, Burghardt, & Weibel, 2008).

This paper presents a study and methodology to define and delineate vaguely defined geographic phenomena. As an example for this class of geographic phenomena, we use city centres, more specifically UK city centres that are delineated from large-scale topographic vector data. There are several reasons for choosing city centres as the example that drives our research. Apart from being an exemplar of a geographic phenomenon whose definition invariably remains vague and is influenced by subjective judgment, city centres are of key functional importance. The city centre, described as the "heart of the city" by Murphy and Vance (1954), functions as nucleus of both business and community activities within the city. For instance, in the context of city centre regeneration, numerous studies investigated topics such as retail development (e.g. Lowe, 2005; Thurstain-Goodwin & Unwin, 2000), visitor activity patterns (Bromley, Tallon, & Thomas, 2003), community safety (Townshead & Pain, 2000), and city centre access and pedestrian movement (Borgers & Timmermans, 1986).

City centres may also serve as an example of how spatial patterns are used in the map generalisation process. In the classical version of *cartographic generalisation*, city centres are depicted and generalised differently than other urban and suburban areas (SSC, 2005). In *model generalisation*, which includes operations to abstract, aggregate, re-classify and reduce representations in a spatial database (Weibel, 1997), a number of techniques have been proposed based on taxonomies. Their application is however restricted to small changes in representation (Chaudhry & Mackaness, 2008a). Achieving more drastic abstractions requires that the semantics of the phenomena is modelled in a prototypical sense (Nyerges, 1991; Mackaness, 2006). Humans seem to define categories in terms of prototypes that contain the most representative attributes within that category (Rosch, 1978; Mennis, Peuquet, & Qian, 2000). Categories have a graded internal structure, that is, some objects are more typical instances of a category than others.

Acquiring prototypical definitions is challenging for phenomena that are only vaguely defined, such as a city centre. The key aim of this paper therefore is to establish a *user-driven* *methodology* to capture models of higher order geographic phenomena. We conduct an online participant experiment to capture the prototypical meaning of a city centre, and present a procedure to delineate city centres from topographic data. As evaluating the recognition of vaguely defined phenomena is by definition non-trivial, we explore three different methods to evaluate the model and the delineation procedure. The present paper extends on research developed in previous papers (Lüscher, Weibel, & Mackaness, 2008; Lüscher et al., 2009), all pursuing the objective of enriching common, originally cartography-oriented spatial databases with high level semantics. The research questions we aim to address in this paper are as follows:

1. With city centres as an example, how can empirical knowledge be formalised to delineate higher order phenomena from topographic databases?

2. What are possible methods for evaluating the results of the delineation process, and how do they perform?

2 Related work

City centres: An early method to delineate central business districts was proposed by Murphy and Vance (1954). For each urban block, the amount of floor space devoted to retailing and commercial activities were used to compute indices of central business activity. A similar approach to delineate town centres was presented by Thurstain-Goodwin and Unwin (2000), aiming at monitoring of urban retail activities. Employment and floor space data was used to create continuous surfaces of town-centric activity. Montello, Goodchild, Gottsegen and Fohl (2003) conducted experiments in delineating 'downtown' by asking people in the street to draw an outline on a paper map. More recently, crowd-sourcing methods were investigated to delineate vernacular areas. Hollenstein and Purves (2010) used georeferenced images from flickr.com to investigate the vernacular use of city core terms.

Pattern recognition from topographic data: Specialised techniques exist for the recognition of urban structures and patterns, using geometric algorithms and/or statistical methods. Many of these techniques focus on the key feature classes defining the urban environment, roads (e.g. Heinzle & Anders, 2007) and buildings (e.g. Regnauld, 2001), and were originally devised to optimise cartographic quality in map generalisation.

A number of techniques were elaborated to abstract topographic datasets to higher order representations. Most of these make use of morphological variables only. For example, approaches exist to delineate settlement boundaries, based on built density and building footprint area (Joubran & Gabay, 2000; Boffet, 2001; Chaudhry and Mackaness, 2008a). Graph-based measures and building morphology were also used to separate areas of urban land use and period of construction (e.g. Barr, Barnsley & Steel, 2004; Steiniger et al., 2008). Boffet (2001) aggregated urban blocks into districts by means of land use and morphology. She also proposed the use of built density and building footprint area as a means to isolate city centres. However, she did not attempt a systematic study.

There have been proposals in the literature to explicitly model geographic phenomena to improve transparency and expressiveness of the model generalisation process. This means to model the semantics of geographic phenomena as sets of properties and (spatial) relations to other concepts. Mallenby (2007) used such an approach for detecting water features. Thomson (2009) presented a method to separate knowledge from pattern recognition algorithms by ontological reasoning on building types and land use categories. Previous work involving the authors has also successfully exploited the use of ontologies in detecting urban house types (Lüscher et al., 2008), including reasoning in the presence of vagueness (Lüscher et al., 2009).

Acquiring and modelling geographic phenomena is challenging, particularly if the phenomenon is only vaguely defined. Conceptualising geographic phenomena as they are understood and used by people, however, would make the derived representations more useful for many applications as discussed in Section 1. Hence, this paper explores the use of participant experiments to capture semantics and subsequently formalises this empirical knowledge for model generalisation. A second aim of the paper is to develop a procedure to spatially delineate city centres from topographic databases.

3 A method to delineate city centres

3.1 Overview

Figure 1 shows an overview of the proposed procedure of computing and evaluating a city centre. The datasets used for the experiment are introduced in Section 3.2. To gain a solid basis for the physical and functional characteristics that constitute a British city centre according to a broad group of people, a participant experiment was carried out (Section 3.3). Based on the analysis of the participant experiment a model of city centre typicality was established which was used to compute city centre typicality values at each point of a regular raster (Section 3.4). The city centre model consists of (groups of) features that are typical or untypical of a city centre and hence have a positive or negative influence on perceived city centre typicality. For each group of features a separate individual city centre typicality surface is computed. The individual typicality surfaces are finally combined to a single city centre typicality value by weighted summation. A crisp city centre area is obtained by applying region growing and a threshold to the continuous city centre typicality surface (Section 3.5). Finally, we suggest several ways to evaluate the plausibility of the computed city centre areas (Section 3.6).

>> Please place Figure 1 here <<

3.2 Datasets

The following datasets provided by the Ordnance Survey of Great Britain (OS) were used. All are vector datasets.

OS MasterMap® Topography Layer roughly represents what can be seen on a topographic map. The granularity corresponds to a scale of 1:1,250 in urban areas. The area features in the Topography Layer form an exhaustive representation of land cover. The primary classification of the features is provided by an attribute that assigns each feature to one of currently 21 groups, such as *Building*, *Natural Environment*, and *Road or Track*.

OS Points of Interest (POI) is the main dataset used for obtaining functional information. It covers commercial addresses and features of interest classified into a three-level taxonomy. The topmost level encompasses 9 classes: *Accomodation*; *Eating and drinking*; *Attractions*; *Commercial services*; *Sport and entertainment*; *Education and health*; *Public infrastructure*; *Retail*; *Manufacturing and production*; and *Transport*. The most detailed level contains more than 600 classes.

OS MasterMap® Address Layer 2 is another Points of Interest dataset offered by the Ordnance Survey. In comparison to the OS Points of Interest dataset, it additionally encompasses residential addresses. However, our evaluation revealed that the coverage of commercial establishments is rather bad. Hence, both datasets were fusioned into a single functional features dataset, taking residential features from Address Layer 2, and all other features from OS Points of Interest.

OS Strategi® is a vector representation of Ordnance Survey's 1:250,000 scale maps. It hence encompasses many features commonly portrayed on regional scale topographic maps, such as roads and railways, lakes and watercourses, woodlands, and urban areas. Of these, only the extents of urban areas were used in this work. They provided the boundaries of the investigation area for each city.

City status in Britain does not imply a certain population size or that the city's formal boundary encompasses an urban area entirely. For the study, 10 British cities (out of approximately 70 with official city status) were selected that either encompassed an urban area or constituted a distinct main core of a larger urban area: Birmingham, Bristol, Cardiff, Leeds, Liverpool, Glasgow, Manchester, Nottingham, Sheffield, and York. All selected cities have a population of 200,000 and above (ranging from 198,800 for York to 1,028,700 for Birmingham;

Office for National Statistics, 2010), and they all serve as commercial and cultural hubs within an urban neighbourhood. Apart from these commonalities, the cities were chosen to reflect variation in topographic characteristics (e.g. seaside cities, riverside cities, and inland cities), and built structure of the inner city (e.g. densely built vs. inner cities with open space).

3.3 Participant experiment

An experiment in the form of a web-based questionnaire was developed to elicit a prototypical model of a city centre from a broad range of people. The results of the questionnaire were then used to build a model of city centre typicality (or 'city centreness', Section 3.4). Additionally, a part of the questionnaire was used to verify the model output (Sections 3.6.3 and 4.3).

3.3.1 Participants

Participants were recruited in two ways. Firstly, an invitation email was sent to several British academics for distribution among their peers and students. Secondly, the link was published in the bulletin boards of two websites that focus on urban planning and geography (www.skyscrapercity.com and www.geograph.org.uk). To provide an incentive, three book vouchers of £50 each were drawn amongst all participants. In the course of a month (March 2010), 101 completed and valid questionnaires were obtained this way.

70.3% of the respondents were male. Similarly, persons in the age group of 20–29 (36.6% of respondents) and 30–39 (20.8% of respondents), respectively, are somewhat overrepresented (Figure 2).

>> Please place Figure 2 here <<

Participants were also asked to indicate current and former places of residence. 40.6% of the respondents always lived at the same place; 36.6% moved, but always within the UK. 82.2% had been living in the UK for longer than 10 years. The geographic distribution of the respondents (Figure 3) shows peaks where the participating academic institutions are located, but the respondents are reasonably well scattered across the United Kingdom. 14.9% of the places of residence are rural areas, and 85.1% urban areas. 70.3% of the respondents indicated a place of residence that has city status.

>> Please place Figure 3 here <<

3.3.2 Design and procedure

The questionnaire was organised into three parts which the participants had to answer in a fixed order. The full questionnaire is provided as electronic supplementary material and can be downloaded from the journal's website. In the following, the two relevant experiments for defining properties of a city centre are presented and discussed. The third experiment was to rate city centre similarity of scenes from panoramic images. It was used for evaluation and is introduced in Section 3.6.3.

The first part of the questionnaire was meant to capture an uninfluenced, individual image of a city centre. It contained experiments where participants had to describe separately frequent activities, important facilities and services, and optionally physical characteristics of a city centre. Answers were to be provided as free text. The task was introduced as follows:

Please define, briefly, in what aspects a city centre differs from other areas of a city.

To render the task more concrete, we asked specifically for services and facilities:

Please indicate: What kind of services & facilities do you expect to find there (in comparison to other areas)?

In the second experiment, the participants were presented a list of urban features and asked to decide whether the features were typical of a city centre. The list is a subset of the full OS Points of Interest taxonomy which was compiled by considering experiences made in previous studies on city centre use (e.g. Bromley et al., 2003; Tallon & Bromley, 2004) and features visible on common topographic maps. Answers were possible on an ordinal scale between -2 (very untypical) and +2 (very typical). The instructions for grading typicality read as follows:

> The following lists contain certain types of concepts that are to be found commonly in urban areas. Please indicate the degree to which they are typical for a city centre.
>
> Select **'Very typical'** if:
>
> - You think that the concept is typically only found within a city centre.
>
> - If you think the best location to find many of the concepts is a city centre.
>
> - If you think the concept is very characteristic for a city centre.
>
> Select **'Very untypical'** if you wouldn't expect such a concept in a city centre.
>
> Select **'Can be either'** if you think the concept can be found commonly within a city centre as well as outside of it.
>
> If you are not sure about the meaning of a concept and can't answer a question, select **'Don't know'**.

3.3.3 Analysis of participant experiment

Figure 4 shows the results of the urban feature grading experiment. The numbers in brackets are cross-references to equivalent concepts in Table 1.

>> Please place Figure 4 here <<

The task to enumerate important services and facilities in city centres resulted in lists of items by each participant, such as *"shops, restaurants, chain bars, shopping centres"* and *"cafés and restaurants, shops, lots of bus stops, railway stations"*. Often the participants qualified the items they named. Some respondents wrote for example *"more specialised shops"*, *"more diverse restaurants"*, or *"denser/richer variety of shops"*. However such qualifiers were not used. Rather, the occurrence of each concept (such as *bar* and *restaurant*) was counted. There were in total 50 different concepts named by the participants. Table 1 shows the concepts named by at least 5% of respondents. Some of the less commonly named items are hotels (4.95%), tourist offices (4.95%), markets (3.96%) and post offices (3.96%).

>> Please place Table 1 here <<

3.4 Operationalisation of city centre typicality

3.4.1 City centre typicality surfaces

The two tasks in the questionnaire were analysed in combination to obtain a model of perceived city centre typicality (or 'city centreness'). Based on this analysis, groups of features were composed that influence city centre typicality in a positive or negative way, respectively, as will be explained below. The final list of characteristics is shown in Table 2.

>> Please place Table 2 here <<

For each of the items in Table 2, a separate typicality surface was computed in the form of a regular raster grid (details follow in Sections 3.4.2 to 3.4.4). The individual surfaces were finally aggregated into a city centre typicality surface by weighted summation (Equation 1).

$$\boldsymbol{typicality}_{citycentre} = \boldsymbol{c}_m \times \sum_i \left(\boldsymbol{w}_i \times \boldsymbol{typicality}_i\right) + \boldsymbol{c}_a \quad (1)$$

where $\boldsymbol{c}_m = \dfrac{1}{\sum_i |\boldsymbol{w}_i|}$ and $\boldsymbol{c}_a = \boldsymbol{c}_m \times \left(|\boldsymbol{w}_{ret}| + |\boldsymbol{w}_{ind}| + |\boldsymbol{w}_{res}| + |\boldsymbol{w}_{natural}|\right)$

C_m and c_a are normalisation constants that ensure that $0 \leq \boldsymbol{typicality}_{citycentre} \leq 1$. C_a includes all negative weights of Table 2 for land uses that are non-typical of city centres, including retail parks (w_{ret}), industrial areas (w_{ind}), residential areas (w_{res}), and natural open ground ($w_{natural}$).

The weights w_i of the individual typicality surfaces *typicality*$_i$ were determined in two steps. First, the corresponding urban features were ranked according to the typicality values and frequency counts they had received by respondents (cf. Figure 4 and Table 1). Second, those urban features that were considered typical were assigned positive weights between 0 and 1, whereby more typical features received higher weights; untypical features were given negative weights. For example, theatres and museums were named frequently and indicated as very typical since they are hardly located outside of city centres. Thus, they were assigned a weight of 1. Similarly, city centre typicality is high if there is a high concentration of places for eating out and for shopping for special goods, as these concepts were mentioned frequently and received high typicality values by respondents. Restaurants received rather moderate typicality values in Figure 4, although they were frequently named in Table 1. This might be explained because places to eat and drink occur also outside the city centre, but in lower concentrations, and are hence perceived as less typical. Thus places to eat and drink were assigned a weight of 0.75. Office-based services were indicated as somewhat typical and thus received a weight of 0.5. Features such as castles or hospitals seem not to influence city centre typicality according to Figure 4. Conversely, it was observed in the experiments that industrial and suburban residential areas (i.e. terraced, detached and semi-detached housing) are seen as very untypical for city centres and indeed they often serve as bounding features for a city centre. The high negative weight of -4 assigned to these features cancels out effects of nearby city centre features, such that raster cells within industrial and residential areas always have low city centre typicality values. A similar, but less strongly pronounced negative influence was observed for the amount of open ground.

From the analysis of the participant experiment it became clear that features influence city centre typicality in three different ways:

1. Features such as shops, retail services, and bus stops characterise city centres by their concentration (and sometimes diversity). Hence, a frequency-based typicality surface is estimated by Kernel Density Estimation (KDE, Section 3.4.2).

2. Certain features (e.g. town hall, railway/road terminal, cathedral, central library) occur only once (or few times) in a city, but are nevertheless important features in structuring the urban landscape; hence they are termed ‘landmark-like’. Rather than the density, the distance to such features is relevant (Section 3.4.3).

3. Large urban regions such as residential districts and industry parks cannot be modelled by points alone. Industrial areas, for example, are comprised of many features, such as factories, office buildings, and open surfaces, whereas the POI dataset generally only covers the locations of head offices. Thus, such areas have to be created first by means of specific algorithms. Their influence is measured by their proportion in a circular window around each raster pixel (Section 3.4.4).

The creation of typicality surfaces for each of the three categories is now described.

3.4.2 Modelling of frequency-based characteristics

For individual establishments, a surface was computed using Kernel Density Estimation (KDE). KDE requires two parameters: The bandwidth and the kernel function, which determines the weighting of the points. In our case, we used a quadratic kernel function. While it is reported that the choice of the kernel function has little influence on the results (Lloyd, 2007, p. 184), the selection of bandwidth is more important. A number of data-driven methods exist to estimate bandwidth objectively (Jones, Marron, & Sheather, 1996). We employed a plug-in bandwidth estimator provided by Duong (2007). Taking the geographic distributions of establishments as input, bandwidths for typicality surfaces were estimated for a subset of the cities. Based on these estimates, it was decided to use a single bandwidth of 350 m for all surfaces to improve comparability. Each surface was subsequently normalised, such that 0 = *minimum typicality* within the study area, and 1 = *maximum typicality*.

3.4.3 Modelling of landmark-like features

For each of the landmark concepts a typicality surface was computed as a function of the Euclidean distance to the landmark feature. Landmarks can be considered as anchors of cognitive representations of space (Winter, Tomko, Elias, & Sester, 2008). Landmarks can be differentiated based on prominence, uniqueness, and salience. Global landmarks, such as the landmarks in this study, are used for referencing from larger distances in a city. The normalisation for landmark typicality surfaces thus assumes a maximum distance of 3 km, corresponding to the size of a large city centre (e.g. Liverpool). Cells further away than 3 km receive a typicality of 0, and distances between 0 km and 3 km are linearly scaled to values between 1 and 0. The threshold distance of 3 km has been empirically set with typical British cities in mind. For larger (or smaller) cities and/or for more important landmarks (that would have a greater impact), the radius would have to be adjusted accordingly.

3.4.4 Modelling of area-like characteristics

While natural open ground is coded in the Topography Layer (in the form of natural areas and water), residential neighbourhoods and industrial areas are themselves complex concepts that were derived in a separate procedure beforehand. An approach for reliably extracting suburban residential buildings from topographic data was shown in a previous publication (Lüscher et al., 2009). Chaudhry et al. (2009) presented an approach to extract functional sites (such as airports and hospitals) from topographic data. The approach used here follows the idea of Chaudhry et al. (2009), but in a simplified form as there is no iterative growing involved. The functional features were intersected with buildings from the Topography Layer to enrich buildings with functions. Then, the algorithm proceeded as described in Table 3.

>> Please place Table 3 here <<

Figure 5 illustrates residential areas obtained in this way. Areas of terraced and semi-detached housing are delineated as suburban residential, while the high street area in the centre and the park in the eastern part of the extract are excluded.

A typicality surface for each type of urban district was obtained by computing the portion of the respective land use within a circular window of 250 m radius around the central point of each raster cell. The window size is different to the one used for KDE because all features within the window have a constant weight, while the quadratic kernel weights distant points less than points near the window centre. The window sizes were thus chosen such that the volumes enclosed by the windows are approximately equal.

>> Please place Figure 5 here <<

3.5 Boundary formation

While it is possible to produce a fuzzy city centre region from the typicality surface, it makes more sense to produce crisp boundaries for many applications, such as cartographic visualisation, query processing, and urban planning (Couclelis, 1996). A region growing algorithm was developed for automatically determining the boundaries of a city centre. The algorithm initialises a city centre area with the cell of highest city centre typicality within a study area. The area is then iteratively enlarged by adding the cell of highest typicality among all cells that are adjacent to the current area. The process stops when the collected area reaches a certain threshold. The obtained city centre boundaries are finally generalised by morphological operations (i.e., erosion and dilation of the polygon) (Millward, 2004).

The most critical part of the process is finding an appropriate threshold for stopping the growing process. In our case, a best-fit value of 0.5 was chosen by considering comparative city centres (cf. Section 3.6.1). Figure 6a shows the evolution of city centre typicality during the growing process. A city centre is delineated when the computed city centre typicality drops below 0.5, i.e. its typicality line enters the grey shaded area in Figure 6a. Spikes of increased typicality occur when the growing process captures secondary areas of high centre typicality. Furthermore, it can be seen that the typicality behaves very similarly in all cities and decreases in a power or logarithmic function with increasing area and hence with increasing distance to the point of highest typicality. Figure 6b shows the progression of the algorithm in Bristol.

>> Please place Figure 6 here <<

3.6 Comparative evaluation

It is obviously challenging to evaluate vague geographic phenomena such as city centres as there cannot be definite reference data. Here, we propose three different methods for assessing the produced city centre boundaries.

3.6.1 Comparative city centre representations

We consulted alternative sources for delineating city centre areas and used them as comparative representations to validate the boundaries produced by our approach. To this end, we searched the web manually for representations of the city centre of each city. The search mainly focused on maps which explicitly designated a city centre area, such as tourist maps or bus maps. Furthermore, Wikipedia provides narrative descriptions of the extents of some city centres. These descriptions were interpreted and mapped. For each city, we created between one and four alternative representations in this way. Collectively, these descriptions give us hints about the extent of the *vernacular* city centre, but we prefer to call them *comparative city centres* rather than *reference city centres*, as they are themselves vague interpretations, represent an individual opinion, or are the result of a political compromise and are therefore different from people's conceptualisation of a city centre. For example, city centre designations on tourist maps may be biased due to the focus on sites of interest for visitors, that is, sites of historic or cultural significance. Narrative descriptions on Wikipedia such as *"bounded north by St Pauls and Easton, east by Temple Meads and Redcliffe, and west by Clifton and Canon's Marsh"* (http://en.wikipedia.org/wiki/Bristol_city_centre, retrieved 14.04.2010) are difficult to confine and might even contain contradictory statements.

The number of representations obtained depends on the number of sources found and their agreement. For example, sources of comparative centres for Glasgow all agree on the extent of the city centre; hence there is only one comparative representation. There is more disagreement for Bristol, where four different interpretations of the city centre extent were acquired.

3.6.2 Volunteered geographic information

As mentioned in Section 2, information from the internet can be used as a proxy of people's vernacular geographic knowledge. The procedure used in this work follows Hollenstein and Purves (2010) who used flickr.com as source of information. Flickr.com is a website where people upload images and describe them by means of tags. It is also possible to attach a geographic location to the image. Flickr provides a web API for automatically searching and downloading such information.

Locations of georeferenced images tagged as 'city centre' were downloaded from flickr.com. For each study area, a distribution of image locations was obtained in this way. Relatively few image locations were available for many cities, such that no representative pattern could be deduced. The comparison hence focuses on four cities: Birmingham, for which 213 locations contributed by 58 people were available; Glasgow (325 locations contributed by 61 people); Liverpool (248 locations contributed by 39 people); and Manchester (421 locations contributed by 90 people).

Vague footprints were created from the point distributions by means of kernel density estimation (KDE) as described by Hollenstein and Purves (2010). The area within the 80% volume contour was selected for quantitative evaluation as it seemed to produce the most plausible city centre areas in the four cities.

3.6.3 Rating of panoramic images

A task of the participant experiment consisted of a series of 360° panoramic images showing urban scenes. In total 15 panorama sites were prepared, out of which a respondent had to judge 10 randomly selected sites. The 15 sites were selected to cover a range of different categories of environment. 12 of the images were located in Bristol; additional 3 images were selected from Manchester to provide a more varied coverage of city centre situations. The panorama sites showed rather prototypical vistas. In particular, we avoided situations such as through roads bordered by shops, or streets that are within the city centre, but that are poor on features indicative of a city centre. Such situations are difficult to judge from the images alone.

The participants had to decide on the degree to which the scene conformed to a city centre. Answers were again possible on an ordinal scale between -2 and +2. The instructions for grading city centre similarity read as follows (see electronic supplementary material for an example stimulus):

> Please have a look at the following 360° panorama. You can move around the panorama using the scroll bars at the bottom of the picture.
>
> Your task is to judge if this picture is of a city centre.
>
> How do you estimate the similarity to a city centre of the location depicted on this page (-2 = very unlike a city centre, 2 = completely like a city centre)

We also asked whether the participants recognised the place shown on the image, and if so, to indicate its location as detailed as possible. However, only one site (Spring Gardens in Manchester) was frequently recognised. Details about the 15 panorama sites are included in the electronic supplementary material to this article. In Section 4.3, the typicality values estimated by the participants are compared to the computed typicality values.

4 Results

4.1 Computed city centre boundaries

Figures 7 and 8 show the computed city centres versus the comparative city centres for each city. Table 4 makes a quantitative comparison of the overlap between computed and alternative city centre areas. It shows precision and recall values, and the F_1-score, which is the harmonic mean of precision and recall. In Equations 2 and 3, $a_{computed}$ denotes the city centre area as delimited by the algorithm, $a_{comparative}$ denotes the area of comparative/Flickr city centre representations, and $a_{overlap}$ denotes the area where computed and comparative/Flickr city centres overlap.

$$precision = \frac{a_{overlap}}{a_{computed}} \tag{2}$$

$$recall = \frac{a_{overlap}}{a_{comparative}} \tag{3}$$

$$F_1 - score = 2 \times \frac{precision \times recall}{precision + recall} \tag{4}$$

If there are multiple comparative areas for a city, the comparison is twofold: against the intersection of the comparative areas, which act as a narrow interpretation of the city centre, and against the union of the comparative areas as a loose interpretation of a city centre. Values for intersection and union are equal where there is only one comparative city centre (i.e. Glasgow).

>> Please place Figure 7 here <<

>> Please place Figure 8 here <<

>> Please place Table 4 here <<

Owing to the uncertainties inherent to the comparative representations (as discussed in Section 3.6.1), a discrepancy between computed and comparative city centre is not necessarily due to an error of the computational model. To assess the *plausibility* of delineated city centres, large differences between the two types of representation were examined in more detail.

The city centres agree rather well in most cases. In Cardiff, Cathays Park was not entirely included; it hosts buildings of public administration, higher education and museums, which are arranged around a central square. The computational model omitted the area due to the high proportion of green space and the absence of other city centre functions. In Leeds, the main difference is an open area under redevelopment which was not captured as city centre by the computational model. The computed city centre in Liverpool is smaller than the comparative city centre. The main differences are residential and industrial areas not captured by the computational model. Since both comparative areas were derived from tourist maps, these areas are presumably designated as city centre because they contain sites of historic and touristic interest. Finally, York is an interesting case because the city centre is historically tightly confined by town walls. However, there are residential areas within the walls which were excluded, but an area hosting some cultural and public institutions outside of the wall was included.

Notable discrepancies occur for Birmingham, Glasgow, and Manchester. In Birmingham, the computational model delineated a protuberance that expands the city centre to the north-west. The area visually resembles a city centre up to St. Pauls Square. However, including the part beyond that square is rather questionable since it actually consists of a mix of different uses in mostly low-rise buildings. Such mixed, commercially highly active urban areas are often hard to distinguish from ‘true’ city centre areas based on topographic information only.

4.2 Delineated city centres for Flickr image locations

Figures 9 and 10 show contour lines for computed city centre typicality on the left hand side, and densities of Flickr image locations on the right hand side. Glasgow and Manchester agree well with the distributions of Flickr image locations. In the quantitative comparison in Table 4 they achieve now high F_1-scores of 0.81 and 0.88, respectively. The agreement is better than with the comparative city centres, where only F_1-scores of 0.65 and 0.71 were achieved.

For Birmingham, the main difference to the computed city centre is again the protuberance to the north-west. For Liverpool, the image locations suggest that the city centre extends further to the west and includes the dockland. The prominence of the waterfront area in Flickr could be explained by the sites of touristic interest and the scenery in this particular area.

>> Please place Figure 9 here <<

>> Please place Figure 10 here <<

4.3 Empirical city centre typicality for panorama sites

Figure 11 presents the empirical city centre typicality as it was judged by the participants based on the panoramic images. The sites were categorised into different types of environment as judged by the authors in Figure 11. Numbers in brackets indicate site numbers relating to the site locations provided in the electronic supplementary material.

>> Please place Figure 11 here <<

The respondents' judgment of residential and industrial situations is clearer than that of city centre situations. There is considerable variation of perceived typicality between the different city centre categories. Evaluating the respondents' comments on their judgments, it seems that open space (in particular green space) and low rise buildings (i.e., only two or three storeys high) have a strong negative influence on perceived city centre typicality. Two sites were judged rather ambiguously: Bristol Queen's Square, which is within the city centre, but features some green space, two storey buildings and no visible shops or business; and Bristol

Canon's Way, which is a new near city centre development featuring business, leisure and tourist attractions.

The computed city centre typicality was subsequently compared to the empirical values obtained for the panorama locations. Figure 12 shows a scatter plot of the empirical values and a linear least squares regression line. The squared Pearson correlation coefficient of the regression is $r^2 = 0.916$. Since the same set of respondents were used to elicit the knowledge for the computational model as well as for the empirical judgment of panorama sites, this cross-comparison cannot be seen as an independent validation of the computational results. Nevertheless, it shows how consistent the respondents are in their verbal descriptions of city centre functions and their visual judgment of exemplars. Furthermore, the strong correlation seems to indicate that the key functions of a city centre have been picked up by the computational model.

>> Please place Figure 12 here <<

In Figure 12 there is a cluster at very low empirical city centre typicality and one at high typicality. These clusters correspond to the selection of test sites, which were chosen to be prototypical of non-city centre and city centre situations. There is also notable agreement for the two sites Canon's Way and College Green (marked A_1 and A_2 in Figures 12 and 13), which were considered to be less clearly definable as being within or outside the city centre. Within each cluster, the variability of computed typicality is larger than the one of empirical typicality. Many of these discrepancies can be explained through the fact that the participants' judgment was restricted to those clues that were visible in the panorama, while the algorithm had information about the larger surrounding area.

Figure 13 shows the spatial distribution of city centre typicality values in Bristol. The site marked as *I* in Figures 12 and 13 is an industrial site and was thus judged as very untypical for a city centre by the respondents. But the proximity of the city centre and a high street with shopping facilities (blue stretch to the west of site *I*) leads to an increased computed city centre typicality. The site marked as *C* shows Broadmead, Bristol's city centre shopping district, and was judged as being very typical for a city centre. However, due to the remoteness to landmark buildings (which are concentrated in the cluster south-west of *C*) and other functions than shopping, the algorithm assessed the site as being less typical for a city centre.

>> Please place Figure 13 here <<

5 Discussion

This paper argued that modelling the underlying conceptual structure is critical to enable automatic recognition of higher order phenomena from topographic databases. Conceptualisations are often hidden and tools have to be developed to render them explicit, i.e. to specify and clarify involved concepts and their logical structure (Smith & Mark, 2001). Smith and Mark (2001) and Agarwal (2004) conducted participant experiments to elicit conceptualisations for generic geographic concepts. Thomson (2009) used a questionnaire to find out how people relate land use to landscape character. On a similar theme as Thomson's study, this paper aimed at acquiring a highly detailed conceptual model for a single, exemplar geographic concept to allow its automatic recognition from topographic databases. The first, main contribution of the research is thus a top-down approach to model generalisation that employs participant experiments to obtain structural knowledge, which is subsequently used to drive the pattern recognition process. As a second contribution we demonstrated the utility of taking a functional perspective to map generalisation, which is often merely seen as a visual optimisation process.

Two alternative methods to delineate city centres from topographic maps were proposed in the literature. Boffet's (2001) experiments for defining city centre districts employed built density and building footprint area. Heinzle and Anders (2007) proposed to use a combination of street network patterns for locating city centres, such as ring roads and star road patterns.

Road patterns are highly individual to the history and geographic setting of each city. Also, our preliminary experiments showed that built density and building footprint area alone are insufficient predictors, as industrial and commercial districts often have similar morphology to city centres with respect to these properties.

In the remainder of this section, the research questions posed in Section 1 shall be revisited.

With city centres as an example, how can empirical knowledge be formalised to delineate higher order phenomena from topographic databases?

Two tasks were presented to elicit empirical knowledge about a higher order phenomenon from participants. The first task asked for uninfluenced associations of city centre qualities. The second task provided lists of features as stimuli. Comparison of the results produced by the first task to the set of facilities named in the second task reveals some differences, which demonstrate that the type of stimulus used is critical. For example, restaurants were the second most frequently named typical facility, but received a moderately high typicality, whereas theatres were less frequently named, but received a high typicality. In the latter case we assume that participants omit features that they rarely use, but nevertheless are seen as important defining elements (such as theatres).

The rich information produced by the questionnaire was thus analysed in a qualitative process to distil salient patterns (Ritchie & Lewis, 2003). The qualitative approach taken in the analysis, however, involved making some deliberate decisions when formulating a computational model for city centre typicality. Setting weights of individual typicality surfaces in Table 2 required careful consideration of questionnaire results, but there is some vagueness involved which might influence the outcome of the city centre model. Similarly, while data-driven methods for bandwidth selection were used, the influence radius of landmark-like features was justified by domain knowledge.

Most critical is, however, the choice of the typicality threshold, as the delineated city centre is very sensitive to this threshold. This can be seen in Figures 9 and 10, where small changes in the threshold value lead to a much better agreement (in the case of Birmingham) or worse agreement (Glasgow and Manchester) with Flickr representations. We are therefore investigating methods for setting the threshold individually for each city. For example, suburban residential and industrial areas could be used as a mask to define the approximate extent of a city centre.

This work used large-scale topographic data and POI data as input. Of these, the topographic data contained little attribute information. If more semantics were available in the input data (e.g. floor space, type of building usage), it would have simplified some of the basic pattern recognition steps (e.g. the recognition of residential building types as a basis for the definition of residential areas), but it would not have changed the main parts of the proposed methodology, that is, the formalisation of the city centre model and the city centre delineation procedure.

It has to be noted that conceptualisations generally may be variable among different cultures (Straumann, 2010), and urban structures are no exception (Steiniger et al., 2008). Hence, the model derived in this research is considered valid for British city centres only. However, the proposed methodology could also be applied to cities elsewhere and, with modification, also to the extraction of other vaguely defined geographic phenomena.

Finally, city centres can be bounded crisply at physical discontinuities, such as city walls (see York), water bodies, or major roads. While our model currently does not take account of such barriers, they could be included by modifying city centre typicality in raster cells covering barriers, making them harder to cross.

What are possible methods for evaluating the results of the delineation process, and how do they perform?

As field surveys such as ones suggested by Montello (2003) are costly to conduct on a large scale, three alternative methods were used in combination to assess the plausibility of the produced regions. The vagueness of the phenomenon city centre is evidenced in the large variation between representations from different sources (Figures 7 and 8). Examples are Birmingham, where there is a significant difference between individual comparative representations, and Glasgow, where the comparative representation and the Flickr representation differ considerably. To deal with this fuzzyness, 'core' regions, i.e. the intersection of comparative regions, and 'boundary' regions, the union of comparative regions, were used for quantitative comparison. Similarly, graded visualisations of computed city centre and Flickr representations in Figures 9 and 10 allow to visually compare the internal structure of city centres.

However, there are potential biases in both methods. Comparative representations are produced by a single (or few) person(s) and are hence not necessarily based on a general consensus. Flickr representations can be systematically biased both in terms of the people contributing pictures as well as of spatial coverage (Hollenstein & Purves, 2010). For example, locations of scenic prominence are likely to be overrepresented in Flickr.

As a third evaluation method, participants of the experiment were asked to rate scenes in panoramic images for city centre similarity. It has to be kept in mind that the participants did not have any knowledge about the surrounding area beyond what was visible in the panoramic images. The participants hence did not know about the nearby shopping street or the restaurant on the other side of the building. The participants judged the situations displayed on the panoramic images consistently with the output produced by the computational model. The panoramas used in our experiment, however, were mostly either highly typical or untypical of city centre situations. It would thus be interesting to repeat the experiment with more images from more ambiguous situations such as A_1 and A_2 in Figures 12 and 13. This would be necessary to see how well the city centre boundary can be defined based on panoramic image ratings.

Thus, each of the evaluation methods discussed above has limitations. However, by combining all three methods it is nevertheless possible to make statements about the plausibility of the computed regions. Our approach seems to produce city centres that conform well to the representations derived from alternative sources, and it conforms to participants' perceptions at individual spots. Larger differences occur for Birmingham, where the computational model produced a protuberance that expands the city centre to the north-west produced by the computational model, which seems to be wrong.

6 Conclusions and outlook

Representing the world as it is conceptualised by people is of great importance in many situations when interacting with GIS (Egenhofer & Mark, 1993; Montello et al., 2003; Hollenstein & Purves, 2010). For the example of the 'city centre' concept, this study presented a methodology to capture conceptualisations of vaguely defined geographic phenomena and use this knowledge to drive the cartographic pattern recognition process. The concepts that are thus extracted relate to high level semantics and provide an added value to the traditional topographic data of National Mapping Agencies and other data providers. The discussed approach aids them to adapt their data for applications such as map generalisation, integration of datasets, urban planning, and geographic search. Also, since the type of data used in our approach is widely available, the approach has the potential to be applicable worldwide.

We see three main extensions of the proposed approach in future research. Firstly, the weights were determined through analysis of the questionnaire in our experiments. Previous research (Bromley et al., 2003; Hubbard, 2002; Tallon & Bromley, 2004) revealed dependencies of individual city centre use from social group and age. It could thus make sense to calibrate the city centre model to different user groups in order to better represent their view of a city centre. Secondly, the same experiments should be carried out for cities in other countries in order to find out what differences there are in the conceptualisation of city centres between different regions and cultures. Thirdly, while we represented the city centre as an area, it could also be represented as a point, depending on scale (or better: map purpose). This location would be the cognitively most representative point within the city centre (the 'cognitive centre of gravity'). It would be interesting to investigate whether that point would coincide with the location of highest city centre typicality value, the centroid of the area, or the location of a landmark concept such as the town hall.

Acknowledgements

The authors appreciate the support of the Ordnance Survey of Great Britain for funding part of the research and for provision of data. Many thanks to Ross Purves for assistance in designing the questionnaire and improving the paper. We would like to thank all the people who spent time on distributing and completing the questionnaire.

References

Agarwal, P. (2004). Contested Nature of Place: Knowledge Mapping for Resolving Ontological Distinctions Between Geographical Concepts. In M. J. Egenhofer, C. Freksa, & H. J. Miller (Eds.), *Geographic Information Science, Third International Conference, GIScience 2004* (pp. 1–21). Berlin: Springer.

Barr, S. L., Barnsley, M. J., & Steel, A. (2004). On the separability of urban land-use categories in fine spatial scale land-cover data using structural pattern recognition. *Environment and Planning B*, *31*(3), 397–418.

Boffett, A. (2001). Méthode de création d'information multi-niveaux pour la généralisation cartographique de l'urbain. (Unpublished doctoral dissertation). Marne-la-Vallée, Université de Marne-la-Vallée.

Borgers, A., & Timmermans, H. J. P. (1986). City Centre Entry Points, Store Location Patterns and Pedestrian Route Choice Behaviour: A Microlevel Simulation Model. *Socio-Economic Planning Sciences*, *20*(1), 25–31.

Bromley, R. D. F., Tallon, A. R., & Thomas, C. J. (2003). Disaggregating the space-time layers of city-centre activities and their users. *Environment and Planning A*, *35*(10), 1831–1851.

Chaudhry, O. Z., & Mackaness, W. A. (2008a). Automatic identification of urban settlement boundaries for multiple representation databases. *Computers, Environment and Urban Systems*, *32*(2), 95–109.

Chaudhry, O. Z., & Mackaness, W. A. (2008b). Creating Mountains out of Mole Hills: Automatic Identification of Hills and Ranges Using Morphometric Analysis. *Transactions in GIS*, *12*(5), 567–589.

Chaudhry, O. Z., Mackaness, W. A., & Regnauld, N. (2009). A functional perspective on map generalisation. *Computers, Environment and Urban Systems*, *33*(5), 349–362.

Couclelis, H. (1996). Towards an operational typology of geographic entities with ill-defined boundaries. In P. Burrough & A. Frank (Eds.), *Geographic Objects With Indeterminate Boundaries* (pp. 45–55). London: Taylor & Francis.

Davies, C., Holt, I., Green, J., Harding, J., & Diamond, L. (2009). User Needs and Implications for Modelling Vague Named Places. *Spatial Cognition & Computation*, *9*(3), 174–194.

Davies, C., Wood, L., & Fountain, L. (2005). User-centred GI: hearing the voice of the customer. In *Annual conference of the Association for Geographic Information AGI'05, People, Places and Partnerships*, London. http://www.ordnancesurvey.co.uk/oswebsite/partnerships/research/publications/docs/2005/ClareDavies_etal_geo.pdf. Retrieved 25.01.2011.

Duong, T. (2007). ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R. *Journal of Statistical Software*, *21*(7), 1–16.

Egenhofer, M. J., & Mark, D. M. (1995). Naïve Geography. In A. U. Frank & W. Kuhn (Eds.), *Spatial Information Theory. A Theoretical Basis for GIS. Proceedings International Conference COSIT '95* (pp. 21–23). Berlin: Springer.

Hart, G., & Greenwood, J. (2003). A Component Based Approach to Geo-Ontologies and Geodata Modelling to Enable Data Sharing. In M. Gould, R. Laurini, & S. Coulondre (Eds.), *AGILE 2003, 6th AGILE Conference on Geographic Information Science* (pp. 197–206). Lausanne: Presses polytechniques et universitaires romandes.

Heinzle, F., & Anders, K.-H. (2007). Characterising Space via Pattern Recognition Techniques: Identifying Patterns in Road Networks. In W. A. Mackaness, A. Ruas, & L. T. Sarjakoski (Eds.), *Generalisation of Geographic Information: Cartographic Modelling and Applications* (pp. 233–253). Amsterdam: Elsevier.

Heinzle, F., Kopczynski, M., & Sester, M. (2003). Spatial data interpretation for the intelligent access to spatial information in the internet. In *Proceedings of the Twenty-first International Cartographic Association Conference*, Durban, South Africa. http://icaci.org. Retrieved 07.06.2011.

Hollenstein, L., & Purves, R. (2010). Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science*, No 1 (2010), 21–48.

Hubbard, P. (2002). Screen-shifting: consumption, ‘riskless risks’ and the changing geographies of cinema. *Environment and Planning A*, *34*(7), 1239–1258.

Jones, M. C., Marron, J. S., & Sheather, S. J. (1996). A Brief Survey of Bandwidth Selection for Density Estimation. *Journal of the American Statistical Association*, *91*(433), 401–407.

Joubran, J. & Gabay, Y. A (2000). Method for Construction of 2D Hull for Generalized Cartographic Representation. In D. Fritsch & M. Molenaar (Eds.), *19th Congress of the International Society for Photogrammetry and Remote Sensing.* International Archives of Photogrammetry and Remote Sensing. IAPRS Vol. XXXIII, Part B4 (pp. 417–424).

Lloyd, C. D. (2007). *Local Models for Spatial Analysis*. Boca Raton: Taylor & Francis Group.

Lowe, M. (2005). The regional shopping centre in the inner city: a study of retail-led urban regeneration, *Urban Studies*, *42*(3), 449–470.

Lüscher, P., Weibel, R., & Burghardt, D. (2009). Integrating ontological modelling and Bayesian inference for pattern classification in topographic vector data. *Computers, Environment and Urban Systems*, *33*(5), 363–374.

Lüscher, P., Weibel, R., & Mackaness, W. A. (2008). Where is the Terraced House? On The Use of Ontologies for Recognition of Urban Concepts in Cartographic Databases. In A. Ruas & C. Gold (Eds.), *Headway in Spatial Data Handling. Proceedings of the 13th International Symposium on Spatial Data Handling* (pp. 449–466). Berlin: Springer.

Mackaness, W. A. (2006). Automated Cartography in a Bush of Ghosts. *Cartography and Geographic Information Science*, *33*(4), 245–256.

Mallenby, D. (2007). Handling Vagueness in Ontologies of Geographical Information. (Unpublished doctoral dissertation). Leeds, University of Leeds.

Mennis, J. L., Peuquet, D. J., & Qian, L. (2000). A conceptual framework for incorporating cognitive principles into geographical database representation. *International Journal of Geographical Information Science*, *14*(6), 501–520.

Millward, H. (2004). A Vector-GIS Extension for Generalization of Binary Polygon Patterns. *Cartographica*, *39*(4), 55–64.

Montello, D. R. (2003). Regions in geography: Process and content. In M. Duckham, M. F. Goodchild, & M. F. Worboys (Eds.), *Foundations of Geographic Information Science* (pp. 173–189). London: Taylor & Francis.

Montello, D. R., Goodchild, M. F., Gottsegen, J., & Fohl, P. (2003). Where's Downtown?: Behavioral Methods for Determining Referents of Vague Spatial Queries. *Spatial Cognition & Computation*, *3*(2), 185–204.

Murphy, R. E., & Vance, J. E. Jr. (1954). Delimiting the CBD. *Economic Geography*, *30*(3), 189–222.

Nyerges, T. L. (1991). Representing geographical meaning. In B. P. Buttenfield & R. B. McMaster (Eds.), *Map Generalization. Making Rules for Knowledge Representation* (pp. 59–85). Essex: Longman Scientific & Technical.

Office for National Statistics. (2010). Population estimates for UK, England and Wales, Scotland and Northern Ireland - current datasets. Mid Year Population Estimates 2009. http://www.statistics.gov.uk/statbase/Product.asp?vlnk=15106. Accessed 19.07.2010.

Purves, R. S., Clough, P., Jones, C. B., Arampatzis, A., Bucher, B., Finch, D., … Yang, B. (2007). The design and implementation of SPIRIT: a spatially aware search engine for information retrieval on the Internet. *International Journal of Geographical Information Science*, *21*(7), 717–745.

Regnauld, N. (2001). Contextual Building Typification in Automated Map Generalization. *Algorithmica*, *30*(2), 312–333.

Ritchie, J., & Lewis, J. (Eds.) (2003). *Qualitative research practice: a guide for social science students and researchers.* London: SAGE Publications Ltd.

Rosch, E. (1978). Principles of Categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum.

Smith, B., & Mark, D. M. (2001). Geographical categories: an ontological investigation. *International Journal of Geographical Information Science, 15*(7), 591–612.

SSC (Swiss Society of Cartography). (2005). *Topographic Maps: Map Graphic and Generalization*. Berne (Switzerland): Federal Office of Topography.

Steiniger, S., Lange, T., Burghardt, D., & Weibel, R. (2008). An approach for the classification of urban building structures based on discriminant analysis techniques. *Transactions in GIS*, *12*(1), 31–59.

Straumann, R. K. (2010). Extraction and characterisation of landforms from digital elevation models: Fiat parsing the elevation field. (Unpublished doctoral dissertation). Zurich, University of Zurich.

Tallon, A. R., & Bromley, R. D. F. (2004). Exploring the attractions of city centre living: evidence and policy implications in British cities. *Geoforum*, *35*(6), 771–787.

Thomson, M. K. (2009). Dwelling *on* Ontology – Semantic Reasoning over Topographic Maps. (Unpublished doctoral dissertation). London, University College London.

Thurstain-Goodwin, M., & Unwin, D. (2000). Defining and Delineating the Central Areas of Towns for Statistical Monitoring Using Continuous Surface Representations. *Transactions in GIS*, *4*(4), 305–317.

Townshead, T., & Pain, R. (2000). Community safety in the city centre. *Town and Country Planning*, *69*(4), 120–121.

Weibel, R. (1997). Generalization of Spatial Data – Principles and Selected Algorithms. In M. van Kreveld, J. Nievergelt, T. Roos, & P. Widmayer (Eds.), *Algorithmic Foundations of Geographic Information Systems* (pp. 99–152). Berlin: Springer.

Winter, S., Tomko, M., Elias, B., & Sester, M. (2008). Landmark hierarchies in context. *Environment and Planning B*, *35*(3), 381–398.

Figure 1

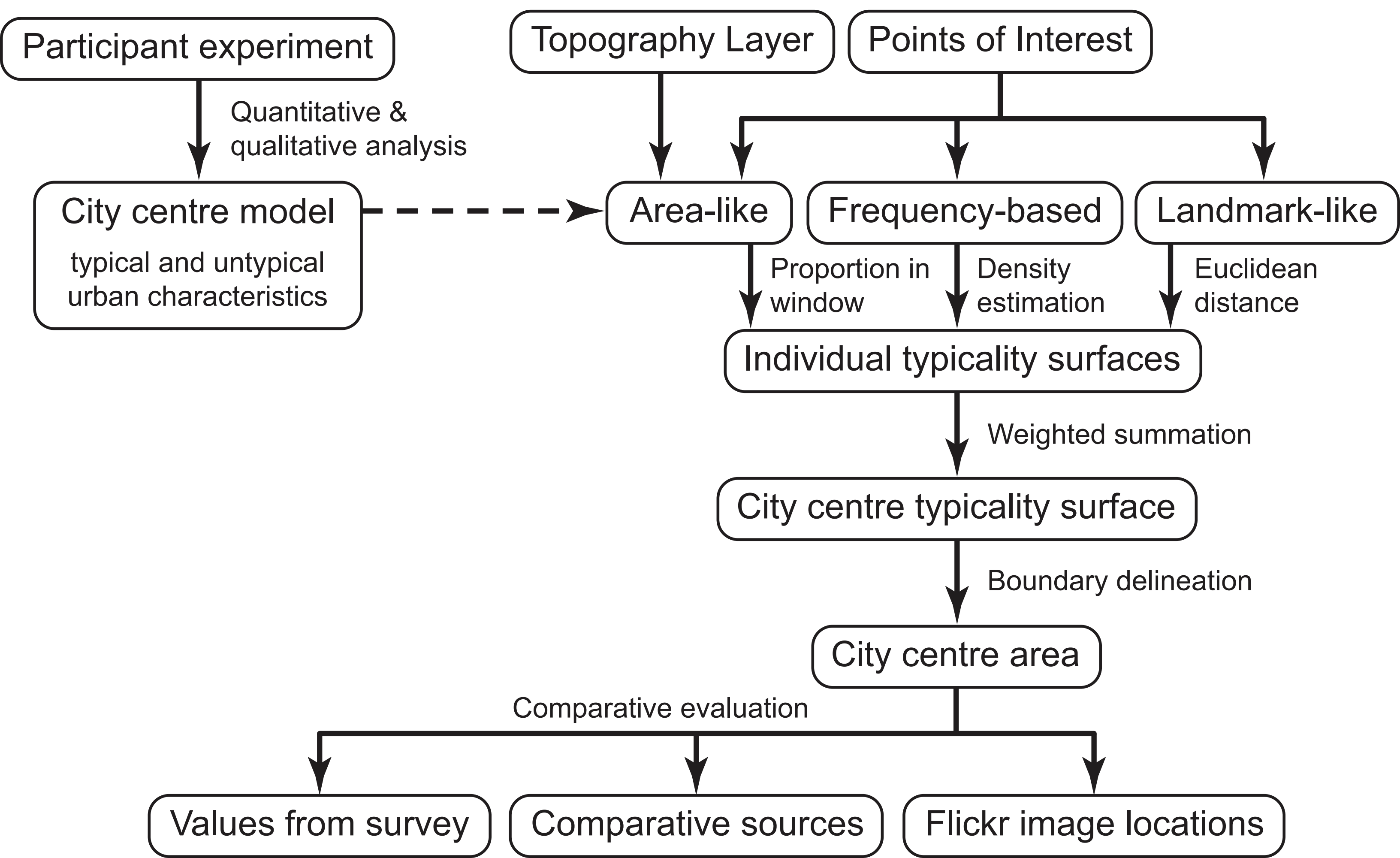
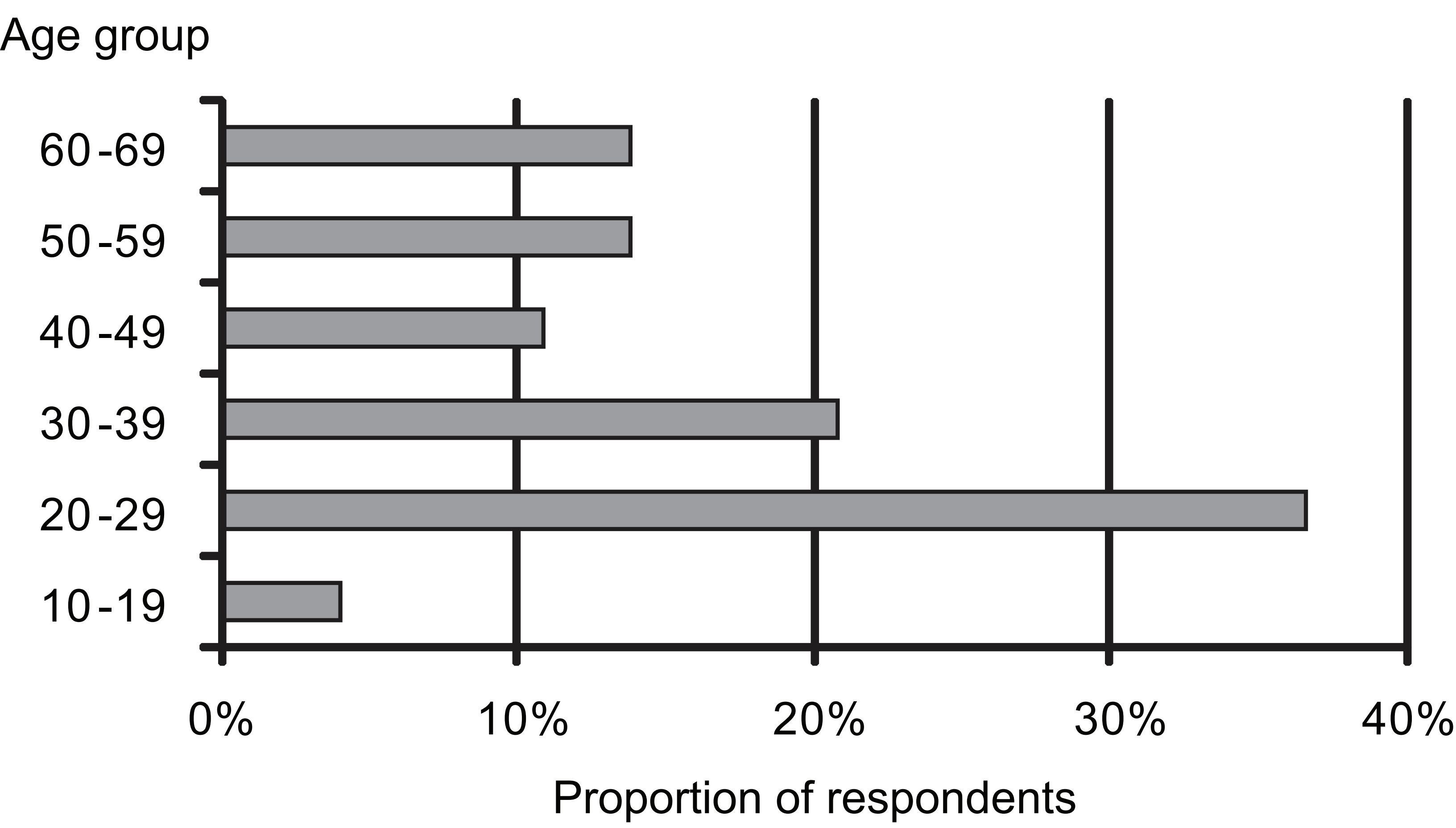

Participant experiment
Quantitative &
qualitative analysis
City centre model
typical and untypical
urban characteristics
Topography Layer
Points of Interest
Area-like
Frequency-based
Landmark-like
Proportion in
window
Density
estimation
Euclidean
distance
Individual typicality surfaces
Weighted summation
City centre typicality surface
Boundary delineation
City centre area
Comparative evaluation
Values from survey
Comparative sources
Flickr image locations

Figure 2



Age group
60-69
50-59
40-49
30-39
20-29
10-19
0%
10%
20%
30%
40%
Proportion of respondents

Figure 3
Click here to download high resolution image

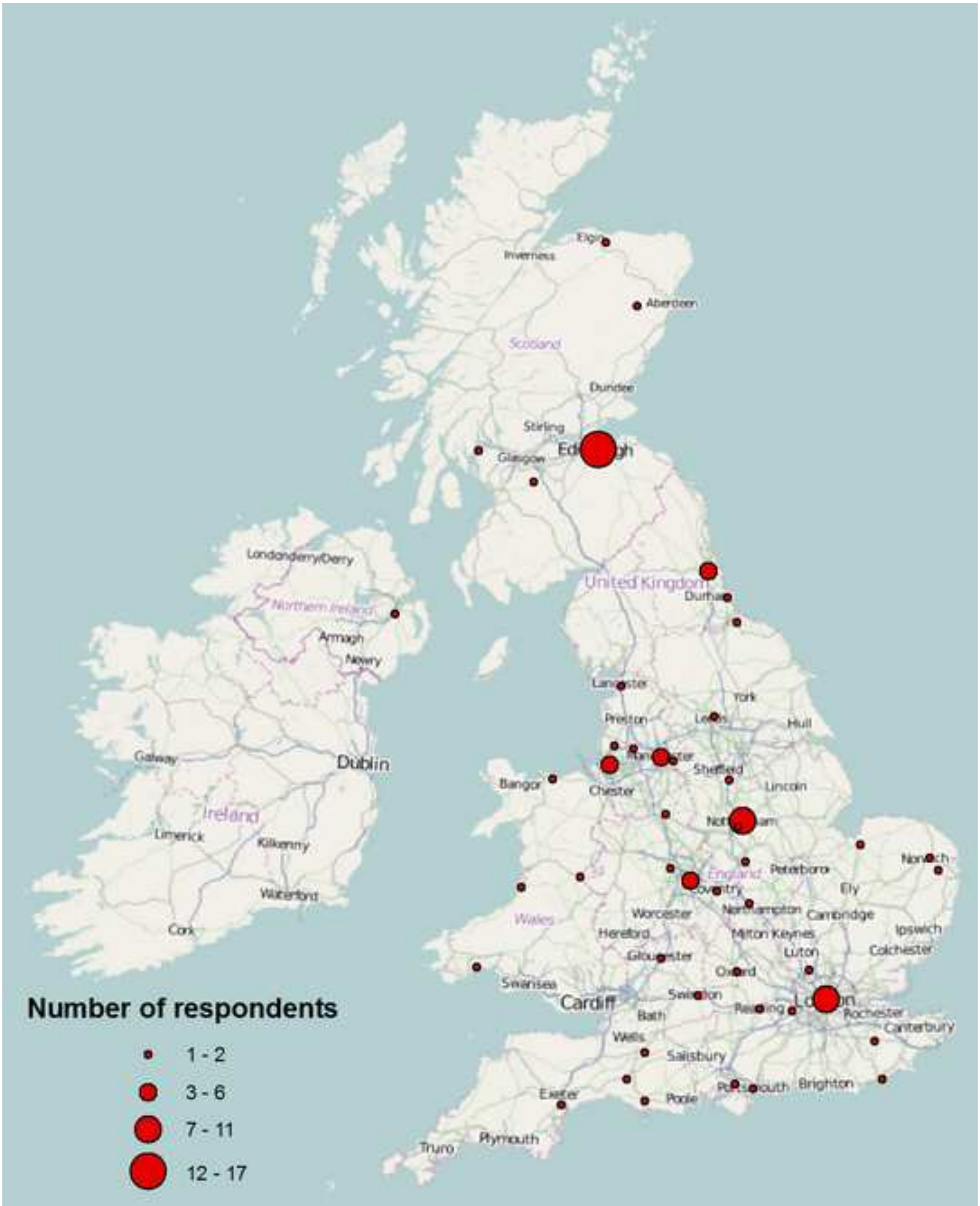
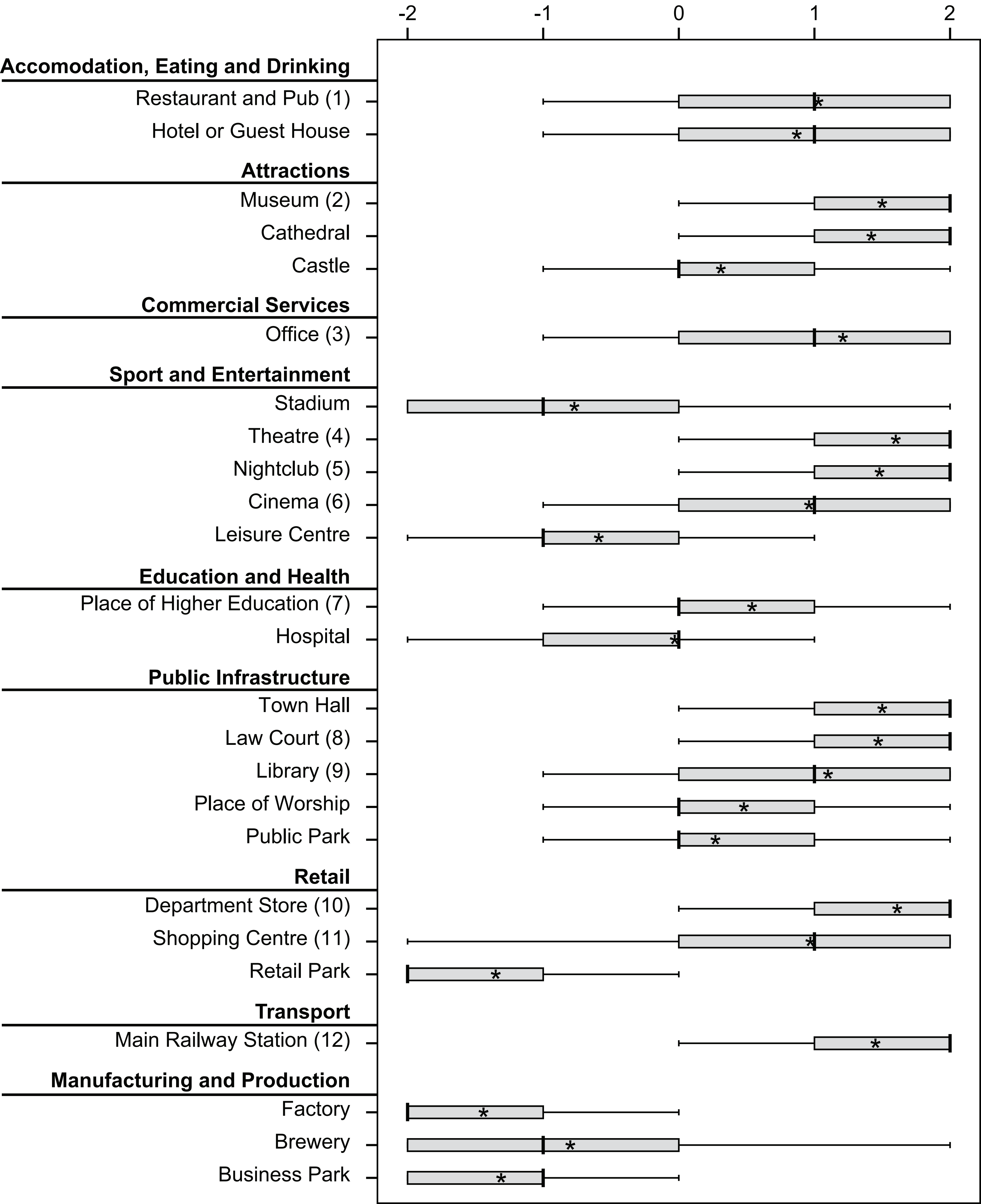


Figure 4



-2
-1
0
1
2
Accomodation, Eating and Drinking
Restaurant and Pub (1)
Hotel or Guest House
Attractions
Museum (2)
Cathedral
Castle
Commercial Services
Office (3)
Sport and Entertainment
Stadium
Theatre (4)
Nightclub (5)
Cinema (6)
Leisure Centre
Education and Health
Place of Higher Education (7)
Hospital
Public Infrastructure
Town Hall
Law Court (8)
Library (9)
Place of Worship
Public Park
Retail
Department Store (10)
Shopping Centre (11)
Retail Park
Transport
Main Railway Station (12)
Manufacturing and Production
Factory
Brewery
Business Park

Figure 5

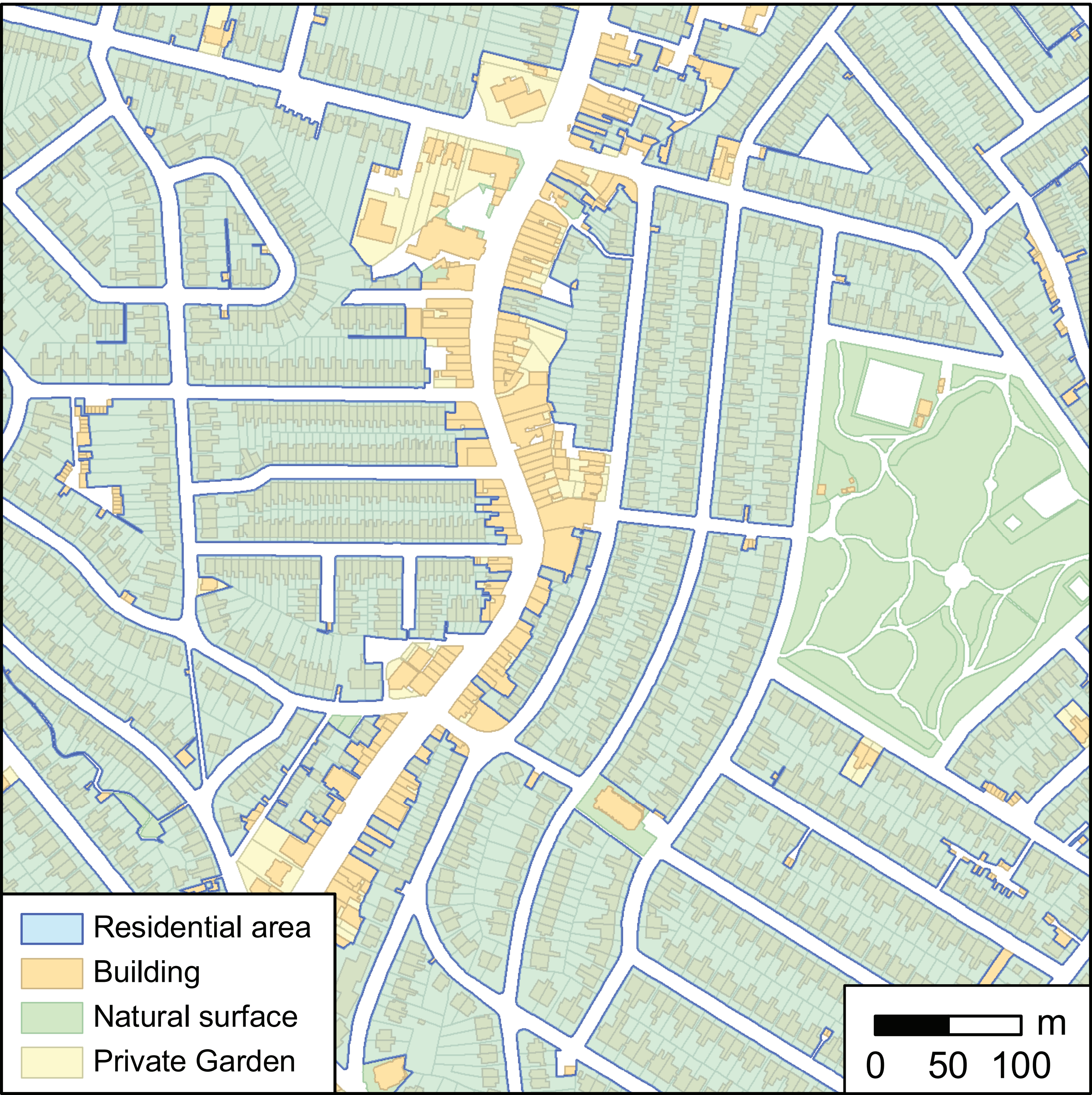
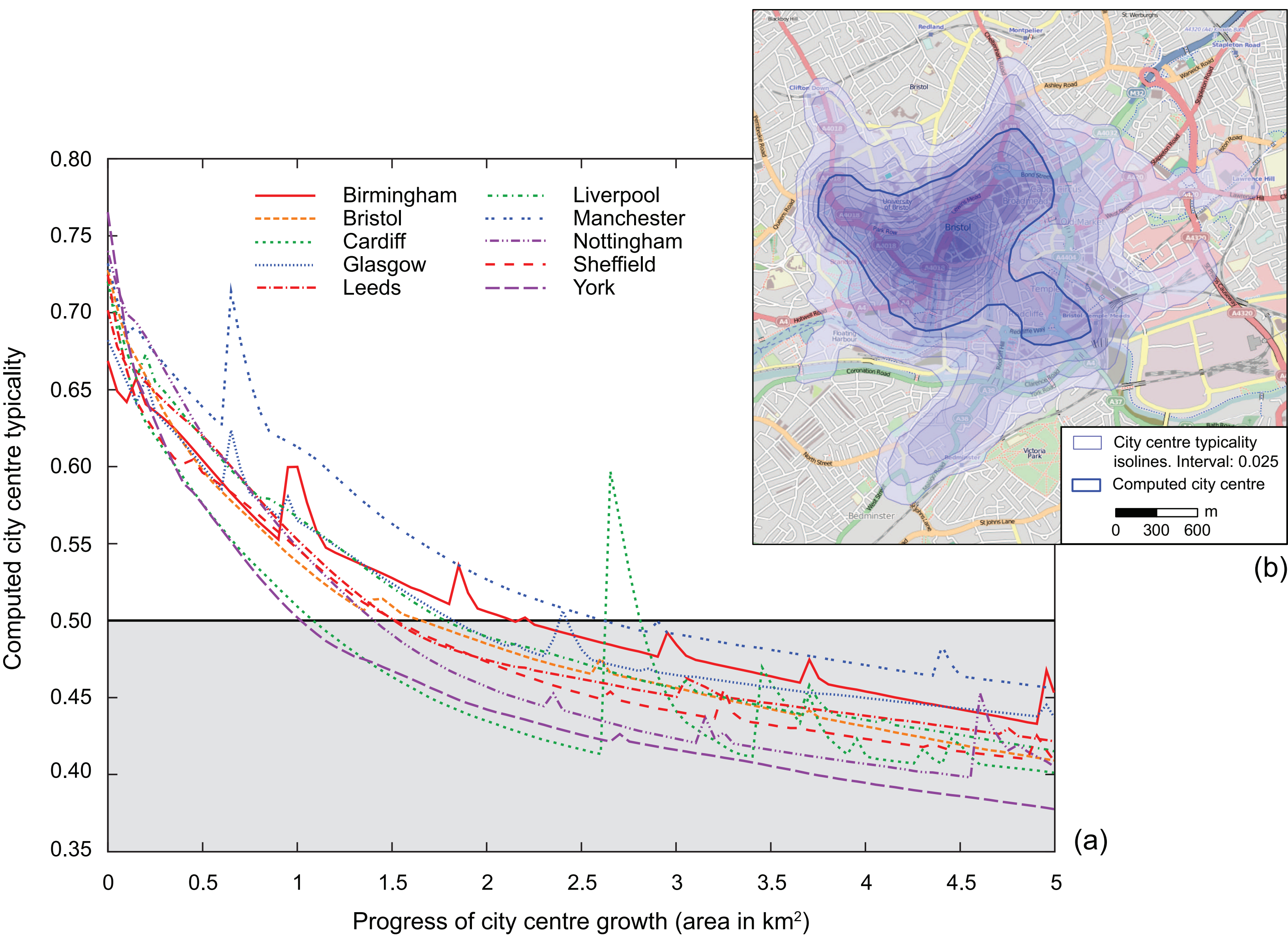

Residential area
Building
Natural surface
Private Garden
m
0 50 100

Figure 6



0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
Computed city centre typicality
Progress of city centre growth (area in km²)
Birmingham
Bristol
Cardiff
Glasgow
Leeds
Liverpool
Manchester
Nottingham
Sheffield
York
(a)
City centre typicality isolines. Interval: 0.025
Computed city centre
m
0 300 600
(b)

Figure 7

Click here to download high resolution image

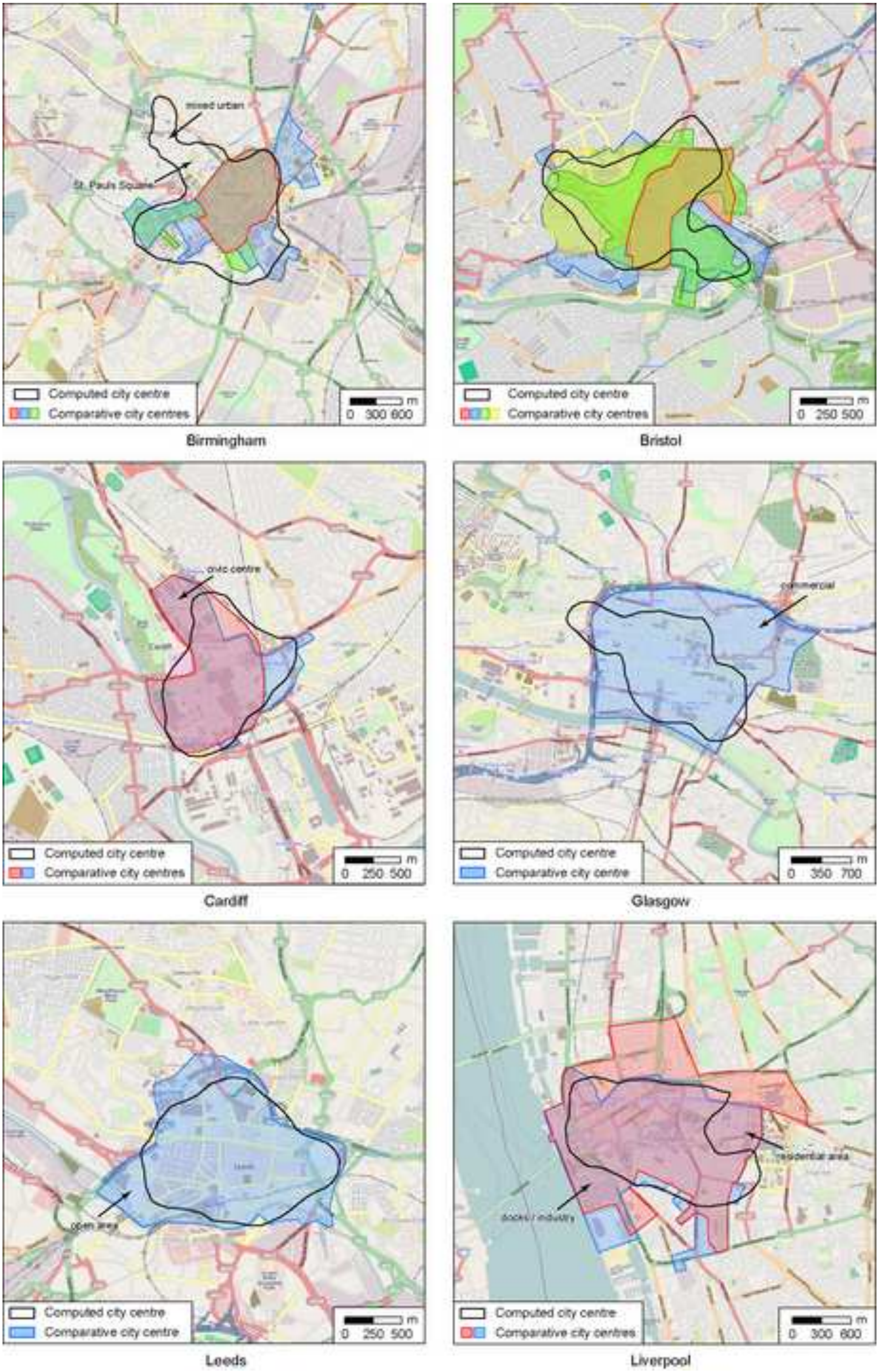


Figure 8

Click here to download high resolution image

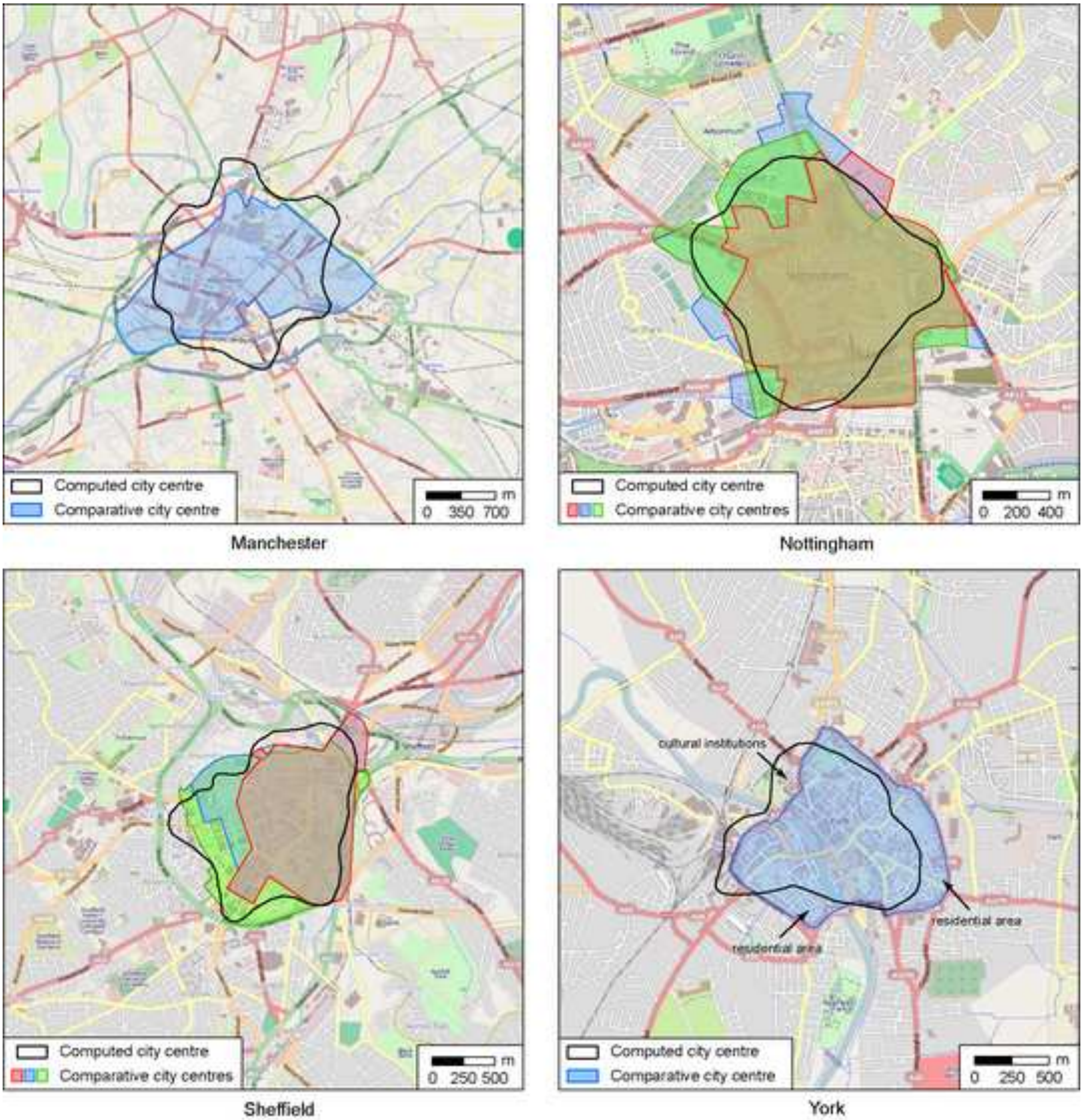
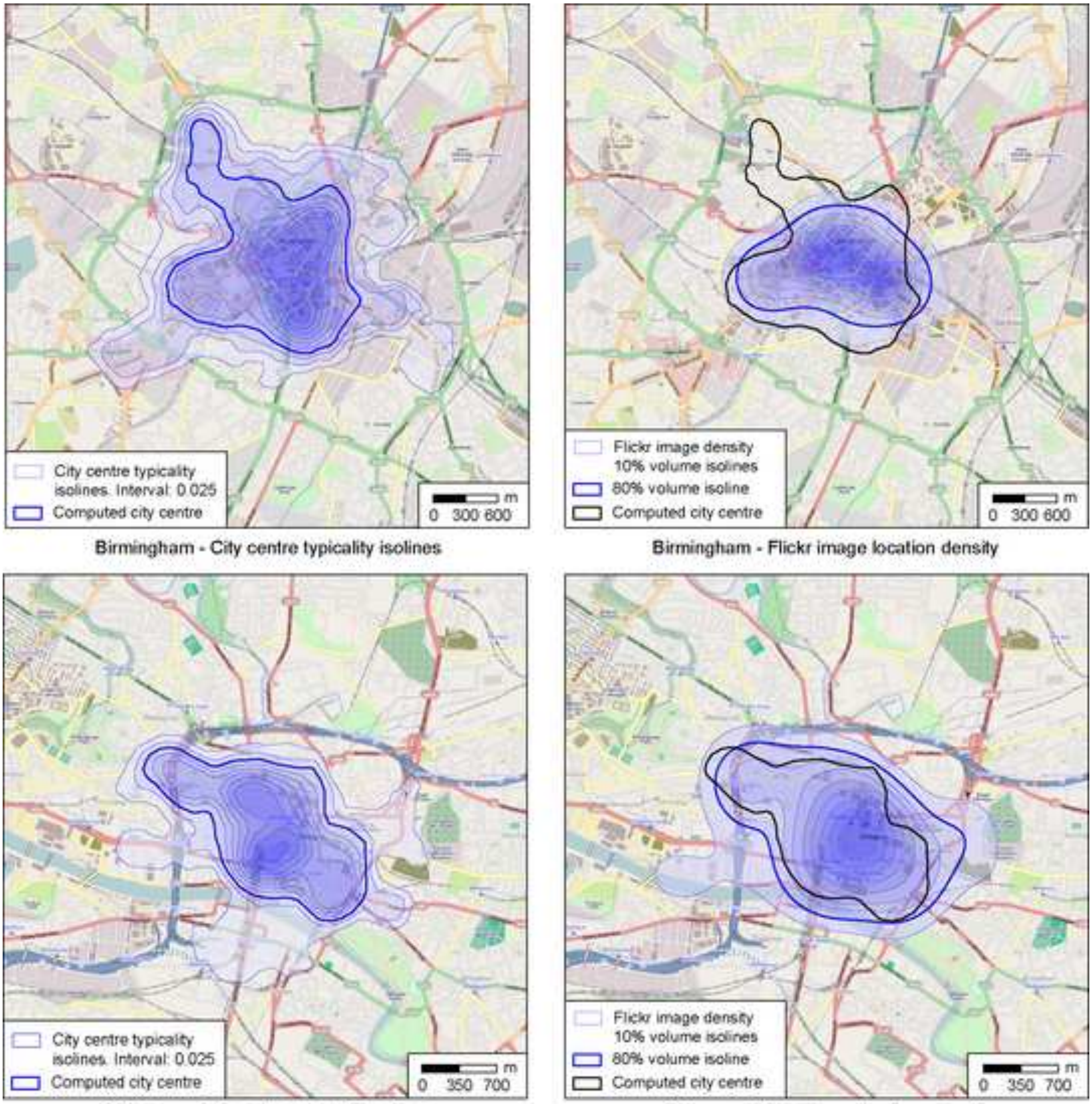


Figure 9
Click here to download high resolution image



Birmingham - City centre typicality isolines

Birmingham - Flickr image location density

Glasgow - City centre typicality isolines

Glasgow - Flickr image location density

Figure 10
Click here to download high resolution image

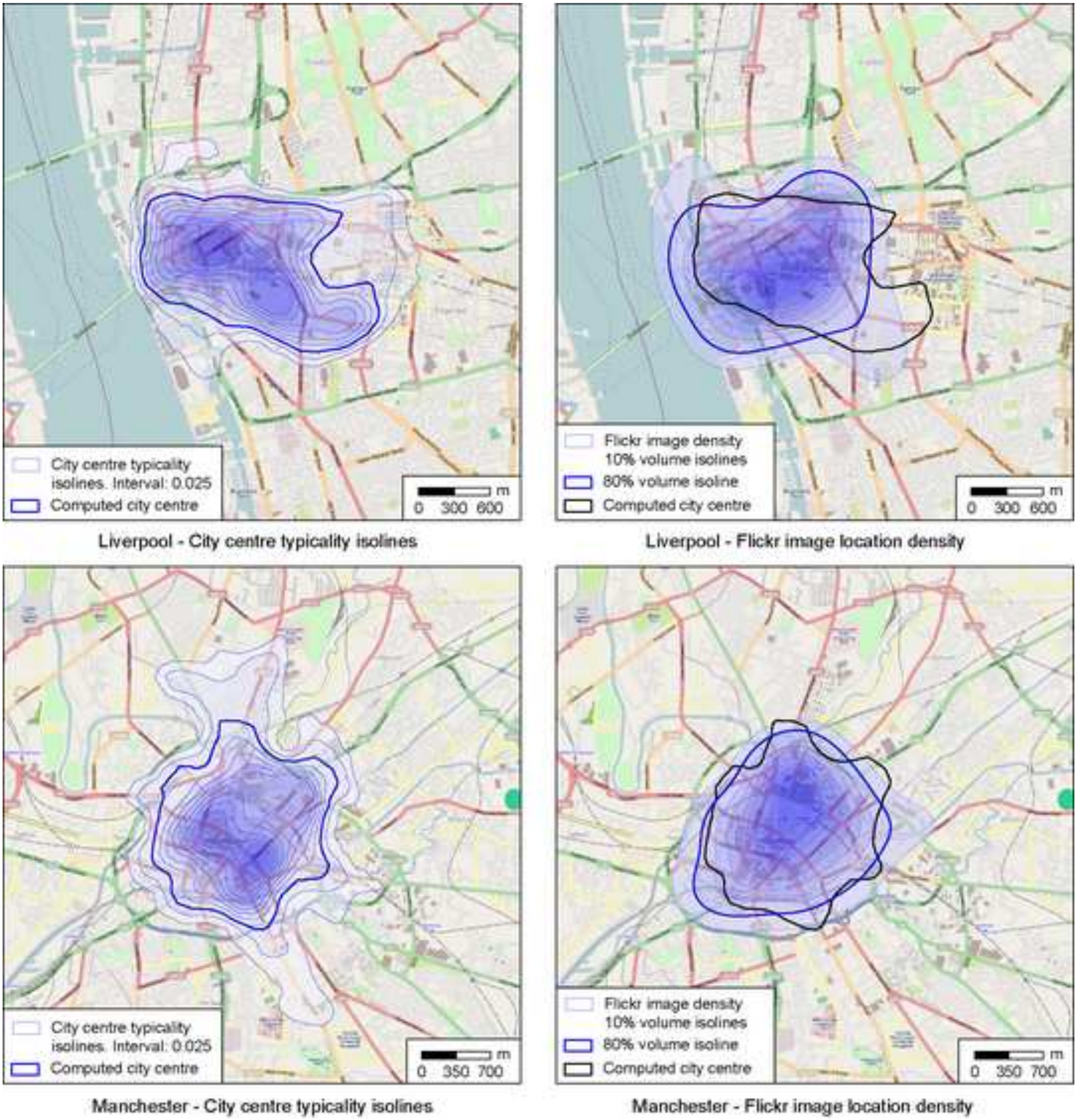
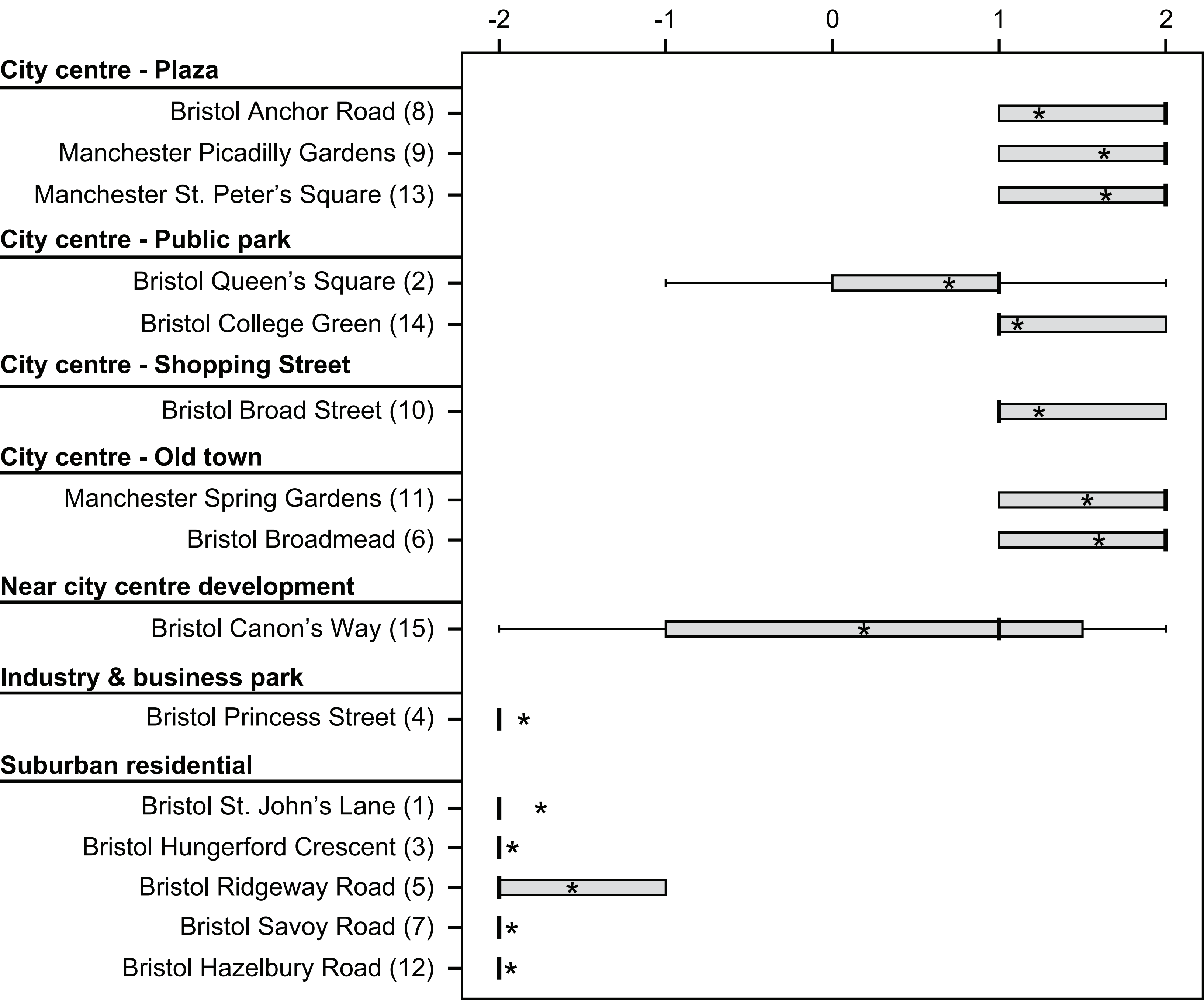
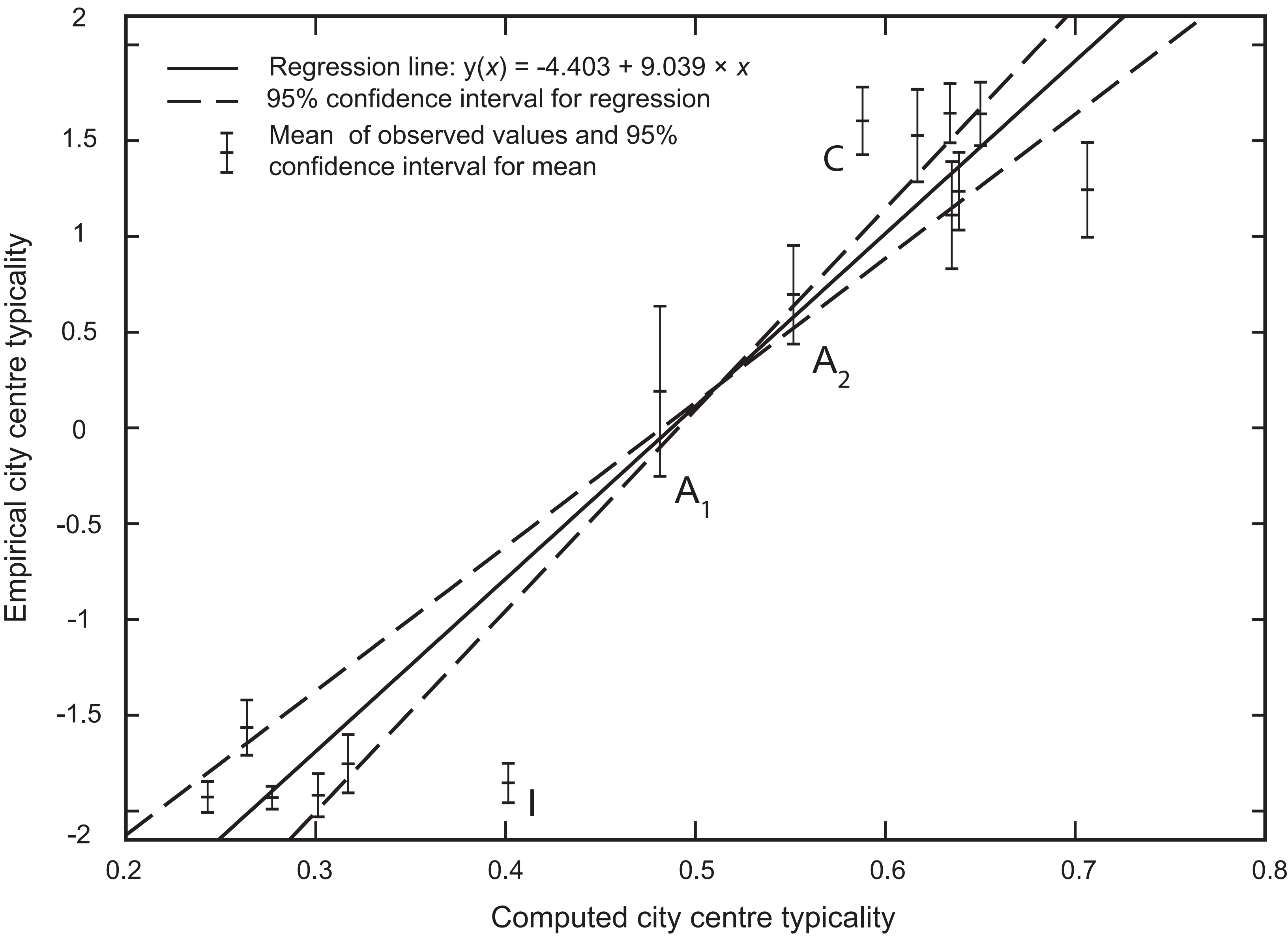


Figure 11



-2
-1
0
1
2
City centre - Plaza
Bristol Anchor Road (8)
Manchester Picadilly Gardens (9)
Manchester St. Peter's Square (13)
City centre - Public park
Bristol Queen's Square (2)
Bristol College Green (14)
City centre - Shopping Street
Bristol Broad Street (10)
City centre - Old town
Manchester Spring Gardens (11)
Bristol Broadmead (6)
Near city centre development
Bristol Canon's Way (15)
Industry & business park
Bristol Princess Street (4)
Suburban residential
Bristol St. John's Lane (1)
Bristol Hungerford Crescent (3)
Bristol Ridgeway Road (5)
Bristol Savoy Road (7)
Bristol Hazelbury Road (12)

Figure 12



Regression line: y(x) = -4.403 + 9.039 × x
95% confidence interval for regression
Mean of observed values and 95% confidence interval for mean
C
A1
A2
I
Empirical city centre typicality
Computed city centre typicality
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0.2
0.3
0.4
0.5
0.6
0.7
0.8

Figure 13
Click here to download high resolution image

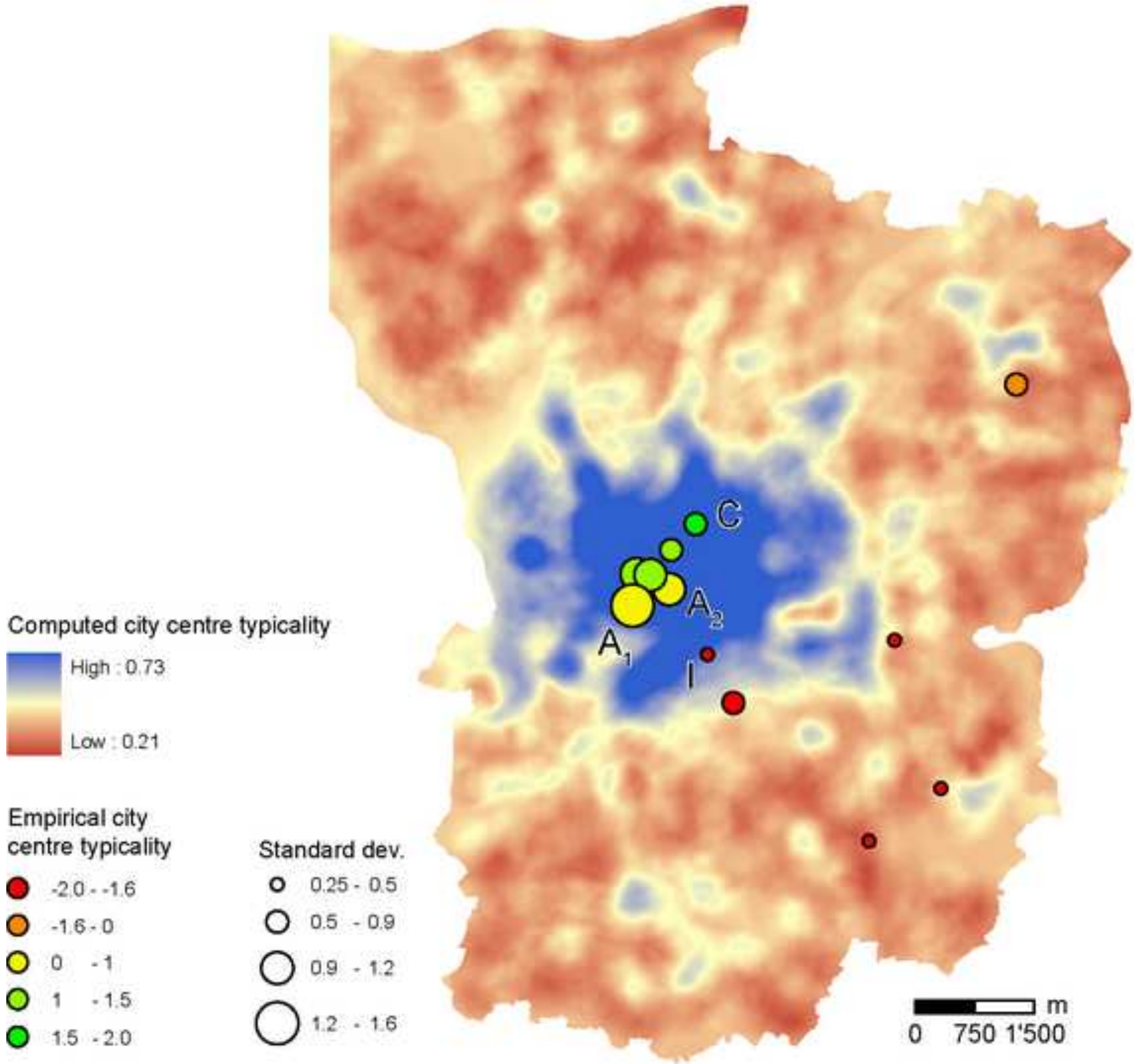


Figure captions

Fig. 1. Overview of procedure for computing a city centre.

Fig. 2. Age structure of respondents.

Fig. 3. Places of residence of respondents. Background mapping © OpenStreetMap contributors, CC-BY-SA.

Fig. 4. Typical (+) and untypical (-) concepts for city centres. The box plots indicate mean (*), median (thick line), and 1st and 3rd quartile (width of boxes). Whiskers include approximately 95% of the responses.

Fig. 5. Extracted suburban residential areas. (Ordnance Survey© Crown Copyright. All rights reserved).

Fig. 6. (a) Plot of city centre typicality against increasing area and (b) contour map of city centre typicality in Bristol. Background mapping © OpenStreetMap contributors, CC-BY-SA.

Fig. 7. Delineated city centres. Background mapping © OpenStreetMap contributors, CC-BY-SA.

Fig. 8. Delineated city centres continued. Background mapping © OpenStreetMap contributors, CC-BY-SA.

Fig. 9. Computed city centre typicality (left) and Flickr image location densities (right) in Birmingham and Glasgow. Background mapping © OpenStreetMap contributors, CC-BY-SA.

Fig. 10. Computed city centre typicality (left) and Flickr image location densities (right) in Liverpool and Manchester. Background mapping © OpenStreetMap contributors, CC-BY-SA.

Fig. 11. Empirical city centre typicality for panoramic image sites. The box plots indicate mean (*), median (thick line), and 1st and 3rd quartile (width of boxes). Whiskers include approximately 95% of the responses.

Fig. 12. Relation between empirical and computed city centre typicality.

Fig. 13. Comparison of spatial distributions of city centre typicality values in Bristol.

Table captions

Table 1. Typical facilities named by the participants.

Table 2. Individual typicality surfaces. Types: F = Frequency-based, L = Landmark-like, A = Area-like.

Table 3. Steps for delineating residential and industrial areas.

Table 4. Comparison of overlap between computed and comparative city centres, and between computed and Flickr delineated city centres.

Table 1

Type of facility	No. of mentions in % of respondents
Accommodation, eating and drinking	
Restaurant (1)	42.57
Pub (1)	24.75
Café	9.91
Attractions	
Museum (2)	21.78
Art Gallery	12.87
Commercial services	
Office (3)	11.88
Sport and entertainment	
Theatre (4)	22.77
Bar	17.82
Night club (5)	12.87
Cinema (6)	10.89
Concert hall / venue	5.94
Education and health	
University (7)	6.93
Public infrastructure	
Civic services & seats of parliament (~8)	20.79
(Main) Library (9)	9.90
Retail	
Shops (boutiques & special goods)	67.33
Bank	13.86
Department store (10)	8.91
Shopping centre (11)	6.93
Manufacturing and production	
None named	
Transport	
Transport hubs (Railway & coach terminals) (~12)	37.62
Dense public transport	21.78

Table 1. Typical facilities named by the participants.

Table 2

Typicality surface	Type	Weight
Accommodation, eating and drinking		
Places to eat and drink (restaurants, pubs, etc.)	F	0.75
Attractions		
Museums and art galleries	F	1
Cathedrals	L	0.5
Commercial services		
Office-based services (stock trading, architects, etc.)	F	0.5
Sport and Entertainment		
Night clubs, amusement arcades	F	1
Theatres, concert halls	F	1
Public infrastructure		
Civic services (consular services, courts, etc.)	F	1
Town hall	L	0.5
Main libraries	L	0.125
Retail		
Boutiques and special goods shops, department stores	F	1
Banks and retail services	F	0.25
Retail parks	F	-1
Transport		
Public transport hubs (main railway stations, coach stations)	L	1
Public transport services (bus stations, tram stations, etc.)	F	0.75
Manufacturing and Production		
Industrial areas	A	-4
Suburban Features		
Suburban residential areas	A	-4
Natural open ground (groves, pastures, bodies of water)	A	-2

Table 2. Individual typicality surfaces. Types: F = Frequency-based, L = Landmark-like, A = Area-like.

Table 3

Residential areas	Industrial areas
Extract residential-only buildings	Extract all buildings that have an industrial function, whereas business services are also allowed
Extract yards that touch the residential buildings	Extract open, manmade and natural surfaces that touch the industrial buildings
Merge residential buildings and yards and dissolve to preliminary residential areas	Merge industrial buildings and open surfaces to preliminary industrial areas
Keep only residential areas that have at least 5 residential buildings	Keep only industrial areas where the portion of industrial building area exceeds 50% of the total building area and that have a total area > 1000 m^2

Table 3. Steps for delineating residential and industrial areas.

Table 4

City	Comparative – Intersection			Comparative – Union			Flickr – 80% Volume Contour		
	Precision	Recall	F_1-score	Precision	Recall	F_1-score	Precision	Recall	F_1-score
Birmingham	0.34	0.97	0.50	0.61	0.77	0.68	0.62	0.90	0.73
Bristol	0.30	0.86	0.45	0.94	0.62	0.75			
Cardiff	0.72	0.78	0.75	0.89	0.77	0.82			
Glasgow	0.92	0.51	0.65	0.92	0.51	0.65	0.98	0.68	0.81
Leeds	0.97	0.71	0.82	0.97	0.71	0.82			
Liverpool	0.88	0.67	0.76	0.96	0.46	0.62	0.79	0.63	0.70
Manchester	0.65	0.77	0.71	0.65	0.77	0.71	0.90	0.87	0.88
Nottingham	0.80	0.84	0.82	0.99	0.68	0.81			
Sheffield	0.62	0.93	0.74	0.91	0.82	0.86			
York	0.85	0.74	0.79	0.85	0.74	0.79			

Table 4. Comparison of overlap between computed and comparative city centres, and between computed and Flickr delineated city centres.